

PCT
 WELTORGANISATION FÜR GEISTIGES EIGENTUM
 Internationales Büro
 INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
 INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



(51) Internationale Patentklassifikation ⁶ : C12N 15/31, 15/52, 15/62, 15/80, 15/81, 1/15, 1/19, C07K 14/37 // C12R 1/75	A1	(11) Internationale Veröffentlichungsnummer: WO 99/31244 (43) Internationales Veröffentlichungsdatum: 24. Juni 1999 (24.06.99)												
<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> <p>(21) Internationales Aktenzeichen: PCT/EP98/07971</p> <p>(22) Internationales Anmeldedatum: 8. Dezember 1998 (08.12.98)</p> <p>(30) Prioritätsdaten: 197 55 302.8 12. Dezember 1997 (12.12.97) DE</p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): HOECHST MARION ROUSSEL DEUTSCHLAND GMBH [DE/DE]; Brüningstrasse 50, D-65929 Frankfurt am Main (DE).</p> <p>(72) Erfinder; und</p> <p>(75) Erfinder/Anmelder (nur für US): KÜCK, Ulrich [DE/DE]; Vorm Felde 11, D-44797 Bochum (DE). SCHMITT, Esther [DE/DE]; Lennershofstrasse 25, D-44801 Bochum (DE).</p> </div> <div style="width: 48%;"> <p>(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Veröffentlicht <i>Mit internationalem Recherchenbericht. Vor Ablauf der für Änderungen der Ansprüche zugelassenen Frist; Veröffentlichung wird wiederholt falls Änderungen eintreffen. Mit Angaben über hinterlegtes biologisches Material, eingereicht gemäss Regel 13bis getrennt von der Beschreibung.</i></p> </div> </div>														
<p>(54) Title: NOVEL TRANSCRIPTION FACTOR CPCRI FOR REGULATION OF ISOPENICILLIN N SYNTHETASE GENE (PCbC) AND AC/ACV SYNTHESASE GENE (pcbAB)</p> <p>(54) Bezeichnung: NEUER TRANSKRIPTIONSFAKTOR CPCRI ZUR REGULATION DES ISOPENICILLIN N-SYNTHETASE-GENS (PCbC) UND DES AC-/ACV-SYNTHETASE-GENS (pcbAB)</p> <p>(57) Abstract</p> <div style="display: flex;"> <div style="width: 35%;"> <p>The invention relates to a DNA binding protein which binds to double-strand DNA of sequence GTTGCCGGGGCCAATCCCTGAGCTT from the intergenic region of the pcbAB and pcbC gene of acremonium chrysogenum as a monomer, homodimer or part of a heterodimer. Said protein can affect the transcription of the pcbAB and/or pcbC gene and is especially characterised in that it contains the amino acid sequence of the CRCR1 protein according to table 7. The invention also relates to fusion proteins containing the amino acid sequence of the CPCRI protein or parts thereof and a transactivation domain. The invention further relates to DNA coding for the known proteins, vectors containing said DNA, host cells containing DNA and/or said vectors, a method for the production of the DNA binding protein, and a method for increasing cephalosporin C yield in acremonium chrysogenum. The invention also relates to DNA binding points for the inventive DNA binding protein.</p> </div> <div style="width: 65%;"> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 15%;"></th> <th style="width: 40%; text-align: center;"> BSII GTTGCCGGGGCCAATCCCTGAGCTT CAACGGCCCGGTTAGGGACTCGAA </th> <th style="width: 45%; text-align: right;"> 603 +/- 165 </th> </tr> </thead> <tbody> <tr> <td style="text-align: right;">BSII m1</td> <td style="text-align: center;"> GTTGCCGGGGGGTTACCCTGAGCTT CAACGGCCCCCAATGGGACTCGAA </td> <td style="text-align: right;"> 9,0 +/- 2,5 </td> </tr> <tr> <td style="text-align: right;">BSII m2</td> <td style="text-align: center;"> GTTGCCGGGGCGTATCCCTGAGCTT CAACGGCCCGCATAGGGACTCGAA </td> <td style="text-align: right;"> 140 +/- 62 </td> </tr> <tr> <td style="text-align: right;">BSII m3</td> <td style="text-align: center;"> CAACCCGGGGCCAATCCCTGAGCTT GTTGGGCCCGGTTAGGGACTCGAA </td> <td style="text-align: right;"> 5,7 +/- 2,0 </td> </tr> </tbody> </table> </div> </div>				BSII GTTGCCGGGGCCAATCCCTGAGCTT CAACGGCCCGGTTAGGGACTCGAA	603 +/- 165	BSII m1	GTTGCCGGGGGGTTACCCTGAGCTT CAACGGCCCCCAATGGGACTCGAA	9,0 +/- 2,5	BSII m2	GTTGCCGGGGCGTATCCCTGAGCTT CAACGGCCCGCATAGGGACTCGAA	140 +/- 62	BSII m3	CAACCCGGGGCCAATCCCTGAGCTT GTTGGGCCCGGTTAGGGACTCGAA	5,7 +/- 2,0
	BSII GTTGCCGGGGCCAATCCCTGAGCTT CAACGGCCCGGTTAGGGACTCGAA	603 +/- 165												
BSII m1	GTTGCCGGGGGGTTACCCTGAGCTT CAACGGCCCCCAATGGGACTCGAA	9,0 +/- 2,5												
BSII m2	GTTGCCGGGGCGTATCCCTGAGCTT CAACGGCCCGCATAGGGACTCGAA	140 +/- 62												
BSII m3	CAACCCGGGGCCAATCCCTGAGCTT GTTGGGCCCGGTTAGGGACTCGAA	5,7 +/- 2,0												

(57) Zusammenfassung

Die Erfindung betrifft ein DNA-bindendes Protein, welches als Monomer, Homodimer oder Teil eines Heterodimer an doppelsträngige DNA der Sequenz GTTGCCGGGCCAATCCCTGAGCTT aus dem intergenischen Bereich des pcbAB und pcbC-Gens von *Acremonium chrysogenum* bindet und in der Lage ist, die Transkription des pcbAB und/oder pcbC-Gens zu beeinflussen, insbesondere dadurch gekennzeichnet, daß es die Aminosäuresequenz des CPCRI-Proteins gemäß Tab. 7 enthält. Des weiteren betrifft die Erfindung Fusionsproteine, enthaltend die Aminosäuresequenz des cpcR1-Proteins oder Teile davon und eine Transaktivierungsdomäne. Die Erfindung betrifft ebenfalls für die genannten Proteine kodierende DNA, Vektoren enthaltend diese DNA, Wirtszellen enthaltend die DNA und/oder die Vektoren, ein Verfahren zur Herstellung des DNA-bindenden Proteins, ein Verfahren zur Steigerung der Cephalosporin C-Ausbeute in *Acremonium chrysogenum*. Außerdem betrifft die Erfindung die DNA-Bindungsstellen für das erfindungsgemäße DNA-bindende Protein.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
	Kuba	LC	St. Lucia	RU	Russische Föderation		
	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
	Deutschland	LK	Sri Lanka	SE	Schweden		
	Dänemark	LR	Liberia	SG	Singapur		
	Island						

Beschreibung

Neuer Transkriptionsfaktor CPCR1 zur Regulation des Isopenicillin N-Synthetase-Gens (*pcbC*) und des AC-/ACV-Synthetase-Gens (*pcbAB*)

Die Erfindung betrifft ein DNA bindendes Protein, welches an doppelsträngige DNA der Sequenz GTTGCCGGGCCAATCCCTGAGCTT (SEQ ID NO.: 4) aus dem intergenischen Bereich des *pcbAB* und *pcbC*-Gens von *Acremonium chrysogenum* bindet und in der Lage ist, die Transkription des *pcbAB* und/oder *pcbC*-Gens zu beeinflussen.

1. Regulation der Genexpression mittels Transkriptionsfaktoren bei Pilzen

Die Genexpression wird u. a. über die Häufigkeit der Transkriptionsinitiation reguliert. Das Vorhandensein von sogenannten Transkriptionsfaktoren kann dabei einen positiven oder einen negativen Einfluß haben. Transkriptionsfaktoren sind *trans*-wirkende DNA-Bindeproteine, die zum Beispiel an DNA-Elemente innerhalb der regulatorischen Sequenzen eines Gens binden können. Die Spezifität der DNA-Protein-Interaktion wird durch Kontakte zwischen den beteiligten Nukleotiden, bzw. dem DNA-Rückgrad und den Aminosäuren und ihren Seitenketten erreicht (Zusammenfassung bei Pabo und Sauer 1992).

Transkriptionsfaktoren werden nach den strukturellen Eigenschaften ihrer DNA-Bindedomänen in verschiedenen Klassen eingeteilt. So kennt man zum Beispiel Helix-Turn-Helix-, Homeodomain-, Zinkfinger oder auch Leucin-Zipper-Proteine (Zusammenfassung bei Pabo und Sauer 1992). Die meisten Klassen von Transkriptionsfaktoren wurden auch bei Hyphenpilzen entdeckt und charakteristische Beispiele werden in Tabelle 1 genannt. Insgesamt wurden etwa 50 Transkriptionsfaktoren von Hyphenpilzen molekular charakterisiert. Mit circa 20 klonierten Transkriptionsfaktoren gehört *Aspergillus nidulans* unter den Hyphenpilzen zu den Organismen, über deren transkriptionelle Regulation der Genexpression am meisten bekannt ist. Bei *Neurospora crassa* liegen detaillierte Kenntnisse zum Beispiel über die Regulation der Gene für die Verwertung von Nitrat

zur Deckung des Stickstoffbedarfs vor (Marzluff 1993). Das *nit-3* Gen kodiert für die Nitratreduktase und stellt das erste Gen des Stoffwechselweges für die Nitrataassimilation dar. Die Expression dieses Gens wird von zwei regulatorischen Proteinen (NIT2 und NIT4) gesteuert. Wobei es sich bei NIT2 um ein allgemeines regulatorisches Protein und bei NIT4 um einen Stoffwechselweg-spezifischen Transkriptionsfaktor handelt (Chiang und Marzluff 1995). NIT4 gehört zur Klasse der $Zn(II)_2Cys_6$ -Zinkfingerproteine, die bislang nur bei Pilzen gefunden wurden (Schjerling und Holmberg 1996). Ein anderes sehr gut analysiertes Beispiel dieser Klasse ist das GAL4-Protein der Hefe *Saccharomyces cerevisiae*.

Aus dem β -Lactam-Produzenten *Acremonium chrysogenum* wurden bisher keine Transkriptionsfaktoren isoliert. Es wurden lediglich biochemische Charakterisierungen von DNA-Bindeproteinen vorgenommen, die mit Cephalosporin C-Biosynthese-Genen in Wechselwirkung treten (Radzio 1996, Then Bergh et al. 1996).

2. Produktion von β -Lactamantibiotika durch Mikroorganismen

Die β -Lactame sind seit der Entdeckung des Penicillins durch Alexander Fleming im Jahre 1929 eine der erfolgreichsten Wirkstoffgruppen mit antibakterieller Potenz. Ursprünglich als eine Kontamination auf einer Petrischale mit Staphylococcen aufgrund seiner wachstumshemmenden Eigenschaften entdeckt, konnte der Pilz *Penicillium notatum* zur ersten biotechnologischen Produktion eines Medikamentes genutzt werden. Chain und Florey isolierten 1940 den Wirkstoff Penicillin und setzten ihn klinisch ein. Das chemische Grundgerüst der Penicilline, die 6-Aminopenicillansäure (6-APA), kann verschiedene Substituenten tragen, was die große Vielfalt der β -Lactame bedingt. Penicilline zeigen sich äußerst wirksam gegen Gram-positive Bakterien wie *Staphylococcus aureus*. Allerdings haben viele Bakterienstämme im Laufe der Zeit Resistenzen entwickelt.

Bei der Entwicklung neuer Penicillin-Derivate wurden z.B. halbsynthetische Antibiotika eingesetzt, bei denen eine fermentativ produzierte Vorstufe, das

Penicillin G, chemisch oder enzymatisch zu 6-APA umgesetzt wird. Diese kann anschließend zu den Wirksubstanzen weiterverarbeitet werden. Zum anderen wurde schon frühzeitig damit begonnen, nach weiteren natürlichen β -Lactamen zu suchen.

1953 konnte aus dem Kulturfiltrat des Hyphenpilzes *Acremonium chrysogenum* (syn. *Cephalosporium acremonium*) neben dem bereits bekannten Penicillin N eine weitere β -Lactam-Stammverbindung, das Cephalosporin C isoliert werden. *A. chrysogenum* zählt wie *Penicillium chrysogenum* zu den filamentös wachsenden Pilzen, die sich morphologisch deutlich von Hefen oder einzelligen Bakterien abgrenzen lassen.

Cephalosporine unterscheiden sich von den Penicillinen durch ein modifiziertes Grundgerüst. Sie werden im Gegensatz zu den Penicillinen außer von Hyphenpilzen auch von Bakterien synthetisiert.

Ähnlich wie bei den Penicillinen werden in der Pharmakologie im wesentlichen halbsynthetische Cephalosporine eingesetzt, die ausgehend von der 7-Aminocephalosporansäure (7-ACS) synthetisiert werden. Die heute verwendete "dritte Generation" der Cephalosporine zeichnet sich durch ein breites Aktivitätsspektrum auch gegen Gram-negative Bakterien, von denen viele Resistenzen gegen Antibiotika zeigen, geringe Humantoxizität und gute Stabilität gegen bakterielle β -Lactamasen aus. *A. chrysogenum* ist bis heute der wichtigste Produzent der entsprechenden Vorstufen. Der Biosyntheseweg und die äußeren Faktoren, die auf die Synthese wirken, sind im Prinzip bekannt. Durch konventionelle Mutageneseverfahren und Auswahl hat man Hochleistungsstämme erzeugt, die Antibiotika in großen Mengen synthetisieren. Produzierte der von Fleming isolierte Stamm noch 0,0012 g Penicillin pro Liter Kultur, so erzielt man in heutigen Prozessen ca. 50 g Penicillin oder über 20 g Cephalosporin C pro Liter Kulturmedium. eine notwendige weitere Verbesserung der Selektivität und Produktivität erfordert jedoch immer bessere Produktionsstämme. Dieses ist nicht

mehr allein durch Zufallsmutagenese zu erreichen, sondern erfordert alternative Methoden. Der Einsatz molekulargenetischer Techniken zur gezielten Beeinflussung der Biosyntheseenzyme erscheint deshalb vielversprechend zu sein.

3. Physikalische Organisation des *pcbAB/pcbC* intergenischen Bereichs von *Acremonium chrysogenum*

Beim Versuch, aus *Streptomyces coelicolor* Gene zu klonieren, die an der Biosynthese des Antibiotikums Actinorhodin beteiligt sind, wurde zunächst festgestellt, daß die gesamte Information, die zur Biosynthese notwendig war, auf einem 33 000 Basenpaare langen DNA-Fragment lag. Nach Transformation dieses Fragmentes in einen anderen Stamm von *Streptomyces*, welcher diese Gene nicht enthielt, konnte dieser Actinorhodin synthetisieren. Dieses war der erste Hinweis auf eine Vielzahl von Antibiotikasynteseenen auf einem relativ kurzen DNA-Abschnitt in einem Prokaryonten. Später konnte gezeigt werden, daß die Kopplung der Biosynthesegene für ein anderes Antibiotikum, das Cephamycin, bei *Streptomyces* soweit geht, daß die Information aus zwei verschiedenen Genen, dem für die Epimerase (*cefD*) und dem für die Expandase (*cefEF*), gemeinsam in RNA umgeschrieben wird. Parallel dazu wurde untersucht, ob in Eukaryonten ähnliche Verhältnisse vorliegen. Experimente mit *Penicillium* zeigten, daß das *penDE*-Gen, das für eine Isopenicillin-N-Acyltransferase kodiert, unmittelbar stromabwärts des *pcbC*-Gens lag. Das entgegengesetzt transkribierte *pcbAB*-Gen ist ebenfalls dem *pcbC*-Gen unmittelbar benachbart. Damit wurde gezeigt, daß die Gene für den Biosyntheseweg der Penicilline als "Cluster" organisiert sind.

In *A. nidulans* findet sich die gleiche Anordnung der Penicillinbiosynthesegene wie bei *P. chrysogenum*. Untersuchungen an *A. chrysogenum* oder beim Actinomyceten *Nocardia Lactamdurans* zeigten, daß hier ebenfalls die für die generellen Schritte der Penicillinbiosynthese verantwortlichen Gene in gleicher Weise organisiert sind. bei den bakteriellen β -Lactamproduzenten, welche die Biosynthese über den Cephamring fortsetzen, liegen die Gene für die darauffolgenden Schritte ebenfalls

nebeneinander, in *Acremonium* dagegeben in einem separaten "Gen-cluster" auf dem Chromosom II. Interessant sind Untersuchungen an Hochleistungsstämmen von *P. chrysogenum* bezüglich ihrer Gendosis. In einem Stamm mit erhöhter Penicillinproduktion liegt das *pcbC*-Gen in mehreren Kopien und im Hochleistungsstamm As-P-78 der gesamte "Cluster" der Gene fünffach vor. Durch die Genvervielfachung kommt es zu einer erhöhten Biosynthese der Genprodukte, da mehrere Kopien gleichzeitig exprimiert werden können.

Das β -Lactamantibiotikum Cephalosporin C wird durch die enzymatische Aktivität mehrerer Proteine in *A. chrysogenum* synthetisiert. Die ersten beiden Syntheseschritte werden durch die AC-/ACV-Synthetase und die Isopenicillin N-Synthetase katalysiert. Beide Enzyme werden durch das *pcbAB*- bzw. das *pcbC*-Gen kodiert, welche im Genom benachbart und divergent angeordnet sind. Die Transkription beginnt im gemeinsamen intergenischen Promotorbereich der beiden Gene.

4. Das Hefe ONE-HYBRID-SYSTEM als Methode zur Isolation von Transkriptionsfaktoren

Das ONE-HYBRID-SYSTEM (Clontech 1995) bietet die Möglichkeit, Interaktionen zwischen DNA-Sequenzen und Proteinen zu untersuchen. Die Analyse der DNA-Protein-Interaktion wird ähnlich wie bei dem analogen TWO-HYBRID-SYSTEM (Protein-Protein-Interaktion, Fields und Song 1989) mit Hilfe eines Reportergens in der Bäckerhefe *Saccharomyces cerevisiae* durchgeführt. Das Verfahren wurde von Wang und Reed (1993) entwickelt und erstmals erfolgreich bei der Klonierung eines olfaktorischen Transkriptionsfaktors eingesetzt.

Sowohl die analysierte DNA-Sequenz als auch die präsumptiven DNA-bindenden Proteine des untersuchten Organismus werden in einen Hefestamm eingebracht (siehe Figur 1). Das DNA-bindende Protein wird als Fusionsprotein mit der transkriptionsaktivierenden Domäne des GAL4-Proteins der Hefe exprimiert. Eine Interaktion des DNA-bindenden Polypeptids mit der analysierten DNA-Sequenz,

welche stromaufwärts eines Reportergens integriert wird, ermöglicht die Transkriptionsinitiation des Reportergens durch die fusionierte GAL4-Domäne. Diese transkriptionsaktivierende Proteindomäne gelangt bei der beschriebenen DNA-Protein-Interaktion in räumliche Nähe des Transkriptionsstartpunktes. Als Reportergene werden das *HIS3*-Gen (in dem Vektor pHISi) und das *lacZ*-Gen (in dem Vektor pLACZi) eingesetzt. Das *HIS3*-Gen ermöglicht die direkte Selektion auf solche Hefezellen, bei denen die beschriebene DNA-Protein-Interaktion stattfindet. Das *lacZ*-Gen erlaubt die Quantifizierung der Reportergenaktivität und damit die Messung der Bindungsaffinität zwischen Protein und analysierter DNA-Sequenz. Neben dem Nachweis der DNA-Protein-Interaktion zwischen zwei bekannten Komponenten kann das ONE-HYBRID-SYSTEM auch zur Isolation bislang unbekannter *trans*-wirkender Faktoren einer DNA-Sequenz eingesetzt werden. Hierbei wird mit einer Expressionsgenbank (cDNA-Bank) des untersuchten Organismus gearbeitet. Die cDNA-Bank wird in dem Hefe-Expressionsvektor pGAD424 angelegt, so daß die klonierten cDNAs zusammen mit der GAL4-Domäne in der Hefe exprimiert werden können. Durch die Selektion auf die Expression von Reportergenen können solche cDNAs identifiziert werden, deren Proteinprodukte in diesem *in vivo*-System eine funktionelle DNA-Bindeaktivität besitzen.

Aus dem 1,2 kb umfassenden Promotor wurde eine DNA-Sequenz von 24 Nukleotiden (Position -447 bis -423) ausgewählt und als Zielelement im ONE-HYBRID-SYSTEM zur Isolation eines spezifischen Transkriptionsfaktors eingesetzt (Figur 2). Die Sequenz enthält eine sogenannte CCAAT-Box, die in den Promotoren und Enhancer-Regionen vieler eukaryontischer Gene gefunden wurde (z. B. Nussinov R 1992). Es existieren mehrere unterschiedliche Transkriptionsfaktoren, die an dieses Motiv binden (z. B. Raymondjean et al. 1988). Z.B. wurde das Am Alpha Binding-Protein (AAB) von *Neurospora crassa* aufgrund seiner Affinität zu einem CCAAT-enthaltendem Promotorfragment kloniert. Durch Mutagenese des CCAAT-Motivs konnte die Spezifität des AAB-Transkriptionsfaktors für diese

Sequenz gezeigt werden (Chen und Kinsey 1995). Bei *Aspergillus nidulans* konnte eine DNA-Protein-Interaktion im Bereich einer solchen CCAAT-Box nachgewiesen werden (Then Bergh et al. 1996).

Die Synthese des Antitubikums in *A. chrysogenum* unterliegt einer Reihe von Faktoren (Kohlenstoffquellen, pH-Wert und Sauerstoffgehalt des Mediums, usw.), deren Umsetzung auf die molekulare Regulation der Genexpression unbekannt ist (Übersicht bei Nosek et al. 1997). Transkriptionsfaktoren mit spezifischen Affinitäten für DNA-Sequenzen des regulatorischen Bereichs der *pcbAB/pcbC*-Gene sind Mittel der molekularen Regulation. Bei Kenntnis und Verfügbarkeit solcher Transkriptionsfaktoren und der entsprechenden Gene erschließen sich neue Möglichkeiten, die Regulation der Cephalosporin C-Biosynthese zu untersuchen und die gewonnenen Erkenntnisse auf die Antibiotika-Produktion zu übertragen.

Die der vorliegenden Erfindung zugrundeliegende Aufgabe ist demgemäß das Auffinden von Genen für Transkriptionsfaktoren, die an der Genregulation der Biosynthesegene von Cephalosporin C beteiligt sind, um letztlich eine Ausbeutesteigerung an Cephalosporin C in entsprechenden Fermentationsprozessen zu erreichen.

Überraschenderweise wurde bei der Lösung dieser Aufgabe mittels des Hefe ONE-HYBRIDSYSTEMS das Gen eines DNA-bindenden Proteins (*cpcR1*-Gen) entdeckt, dessen Gen-Produkt (*CPCR1*) im intergenischen Bereich der *pcbAB* und *pcbC*-Gene von *A. chrysogenum* bindet und der Familie der RFX-proteine (Emery et al., 1996) zuzurechnen ist. Dies ist um so überraschender, als daß man aufgrund des Vorliegens einer CCAAT-Box im zum Screening benutzten DNA-Fragment eigentlich ein Protein aus einer Protein-Familie erwartet hätte, welche aus Mitgliedern besteht, die spezifisch an das Sequenzmotiv "CCAAT-Box" binden. Eine mögliche Erklärung für diesen überraschenden Befund ist die Struktur der erfindungsgemäßen DNA Bindestelle: GTTGCGGGCCAATCCCTGAGCTT (SEQ ID No.: 4). Die Sequenz enthält ein palindromisches Sequenzmotiv, welches wohl die Bindungsstelle für das

CPCR1-Protein darstellt; die CCAAT-Box überlappt lediglich teilweise mit dieser Sequenz.

Ein Gegenstand der Erfindung ist ein DNA-bindendes Protein, welches als Monomer, Homodimer oder Teil eines Heterodimer an doppelsträngige DNA der Sequenz GTTGCCGGGCCAATCCCTGAGCTT (SEQ ID NO.: 4) aus dem intergenischen Bereich des pcbAB und pcbC-Gens von *Acremonium chrysogenum* bindet und in der Lage ist, die Transkription des pcbAB und/oder pcbC-Gens zu beeinflussen, insbesondere dadurch gekennzeichnet, daß es die Aminosäuresequenz des CPCR1-Proteins gemäß Tabelle 7 (SEQ ID NO.: 20) oder Teile davon enthält.

Ein weiterer Gegenstand der Erfindung ist ein Fusionsprotein, enthaltend die Aminosäuresequenz des CPCR1-Proteins gemäß Tabelle 7 (SEQ ID NO.: 21) oder Teile davon und eine Transaktivierungsdomäne, welche in der Lage ist, die Transkription des pcbAB und/oder pcbC-Gens zu beeinflussen.

Insbesondere ist das Fusionsprotein dadurch gekennzeichnet, daß die genannten Teile des CPCR1 Proteins der DNA-Bindungsdomäne (Aminosäuren 180 bis 254 gemäß Tabelle 7) und der Dimerisierungsdomäne (Aminosäuren 502 bis 660 gemäß Tabelle 7) entsprechen und/oder die Transaktivierungsdomäne vom Transkriptionsfaktor GAL4 (Aminosäuren 768 bis 881) stammt.

Ein weiterer Gegenstand der Erfindung ist ein DNA-Molekül, kodierend für ein Protein wie vorgehend beschrieben. Dieses DNA-Molekül ist insbesondere dadurch gekennzeichnet, daß

- (a) es eine DNA-Sequenz gemäß Tabelle 6 (SEQ ID NO.: 20) oder Teile davon enthält;
- (b) es eine DNA-Sequenz enthält, die in der Lage ist, mit dem DNA-Molekül gemäß (a) unter stringenten Bedingungen zu hybridisieren, oder Teile davon;

- (c) es eine DNA-Sequenz enthält, die aufgrund der Degeneriertheit des genetischen Codes von den DNA-Molekülen gemäß (a) und (b) verschieden ist, jedoch die Expression der entsprechend mit den DNA-Molekülen gemäß (a) und (b) exprimierbaren Proteinen ermöglicht, oder Teile davon.

Ein weiterer Gegenstand der Erfindung ist ein Vektor, enthaltend eine DNA wie vorhergehend beschrieben. Insbesondere ist dieser Vektor dadurch gekennzeichnet, daß es ein Vektor geeignet zur Expression in *Acremonium chrysogenum* oder *Saccharomyces cerevisiae* ist.

Bevorzugt ausgewählt wird der Vektor aus der Gruppe enthaltend die Vektoren pGC1 und pGC1 Δ DIM1.

Ein weiterer Gegenstand der Erfindung ist eine Wirtszelle, enthaltend eine oder mehrere Kopien eines Vektors oder einer DNA wie vorhergehend beschrieben, insbesondere dadurch gekennzeichnet, daß sie ausgewählt ist aus der Gruppe enthaltend *Acremonium chrysogenum* und *Saccharomyces cerevisiae*.

Ein weiterer Gegenstand der Erfindung ist ein Verfahren zur Herstellung eines DNA-bindenden Proteins wie vorhergehend beschrieben, dadurch gekennzeichnet, daß eine DNA wie vorhergehend beschrieben mittels eines Vektors wie vorhergehend beschrieben in einer Wirtszelle wie vorhergehend beschrieben exprimiert und isoliert wird.

Ein weiterer Gegenstand der Erfindung ist ein Verfahren zur Steigerung der Cephalosporin C-Ausbeute in *Acremonium chrysogenum*, bei dem die Expression des cpcR1-Genprodukts variiert wird, insbesondere dadurch gekennzeichnet, daß die variierte Expression durch Einbringen einer oder mehrerer Kopien des cpcR1-Gens in den Wirtsorganismus zusätzlich zum endogenen cpcR1-Gen, erreicht wird oder auch dadurch gekennzeichnet, daß die variierte Expression durch Integration einer oder mehrerer zusätzlicher DNA-Bindesequenzen TTGCCGGGCCAAT (SEQ

ID NO.: 1) oder Varianten gleicher Funktion des DNA-bindenden Proteins wie vorhergehend beschrieben, im *pcbAB/pcbC*-Promotors erreicht wird.

Ferner ist Gegenstand der Erfindung eine DNA-Bindungsstelle für ein DNA-bindendes Protein wie vorhergehend beschrieben, ausgewählt aus der Gruppe enthaltend die doppelsträngigen Sequenzen GTTGCCGGGCCAATCCCTGAGCTT (SEQ ID NO.: 3), TTGCCGGGCCAAT (SEQ ID NO.: 1) und GTTGCCGGGCGTATCCCTGAGCTT (SEQ ID NO.: 8).

Ein weiterer Gegenstand der Erfindung ist ein Vektor wie vorhergehend beschrieben, enthaltend eine oder mehrere Kopien der DNA-Bindesequenzen des erfindungsgemäßen DNA-bindenden Proteins wie vorhergehend beschrieben.

Die Erfindung wird nun anhand der Beispiele, Figuren und Tabellen beschrieben, ohne darauf beschränkt zu sein.

Beispiel 1:

Isolation der *cpcR1*-cDNA von *Acremonium chrysogenum* mit Hilfe des ONE-HYBRID-SYSTEMS

Die im vorausgegangenen beschriebene DNA-Sequenz aus dem *pcbAB/pcbC*-Promotorbereich wurde eingesetzt um mittels des ONE-HYBRID-SYSTEMS cDNA-Klone von *A. chrysogenum* zu isolieren, die für DNA-Bindeproteine kodieren. Es wurde grundsätzlich so vorgegangen, wie im "Matchmaker One Hybrid System protocol" von Clontech beschrieben. Es wurde ein Oligonukleotid mit einer Länge von 80 Nukleotiden (nt) synthetisiert, welches die DNA-Sequenz (Position -447 bis -423 im *pcbAB/pcbC*-Promotor) in dreifacher Kopie enthält (Tabelle 3). Am 5'-Ende wurden die Nukleotide AATTC zugefügt, um einen *EcoRI*-Überhang zu erzeugen. Am 3'-Ende besitzt das Oligonukleotid die Nukleotide CCC, die zur Erkennungssequenz des Restriktionsenzym *SmaI* gehören. Zudem wurde ein Oligonukleotid von 76 nt synthetisiert, dessen Sequenz komplementär zum ersten

Oligonukleotid ist. Das 3'-Ende dieses Oligonukleotids schließt mit einem G, so daß bei der Hybridisierung der beiden Oligonukleotide ein doppelsträngiges DNA-Molekül entsteht, dessen 5'-Ende durch einen *EcoRI*-Überhang charakterisiert ist. Dieses doppelsträngige DNA-Molekül (BSII) wurde in die Reportergenplasmide pHISi und pLACZi (Clontech) integriert. Die Plasmide wurden zuvor mit den Restriktionsenzymen *EcoRI* und *SmaI* geschnitten. Die erhaltenen Plasmide pHISi-BSII und pLACZi-BSII (siehe Tabelle 4) wurden mittels DNA-Transformation in das Genom des Rezipientenstammes integriert.

Der resultierende Hefestamm (siehe Tabelle 5) wurde zur Transformationen einer cDNA-Bank von *A. chrysogenum* eingesetzt. Als Ausgangsmaterial für die cDNA-Synthese diente mRNA des *A. chrysogenum*-Stammes A3/2 (Radzio und Kück, 1997; DSM 11878) und für die Konstruktion der cDNA-Bank wurde der Hefevektor pGAD424 (Clontech) verwendet.

Es wurden zwei identische Klone identifiziert, die auch in wiederholten *in vivo*-Experimenten zu einer nachweisbaren DNA-Protein-Interaktion mit dem Zielelement führten. Die DNA-Sequenz der isolierten cDNA von *A. chrysogenum* ist in Tabelle 6 dargestellt. Der isolierte Klon umfaßt ein cDNA-Fragment von 2541 bp und zusätzlich ein PolyA-Ende von ca. 80 Nukleotiden.

Beispiel 2:

Homologie des CPCR1-Proteins zu Transkriptionsfaktoren der RFX-Familie

Die DNA-Sequenz der beschriebenen aus *A. chrysogenum* isolierten cDNA enthält einen offenen Leserahmen für 786 Aminosäuren (Tabelle 7). Ein Computervergleich der Proteinsequenz des *cpcR1*-Gens mit der EMBL-Datenbank ergab Homologien von 28 bis 22 Prozent zu verschiedenen Mitgliedern der Familie der RFX-Proteine. Die höchste Homologie besteht mit 28 % zum SAK1-Protein von *Schizosaccharomyces pombe*. Das humane RFX1-Protein ist von allen Mitgliedern der Familie am besten charakterisiert und besitzt zum CPCR1-Protein eine Homologie von 23,9 % über einen Bereich von 476 Aminosäuren. Die Homologien innerhalb der konservierten Proteindomänen liegen deutlich höher (Abbildung 3 und

Tabellen 8 und 9). Die Transkriptionsfaktoren der RFX-Familie sind durch eine Reihe von konservierten Proteindomänen gekennzeichnet (Übersicht bei Emery et al. 1996). Die Proteine besitzen eine 76 Aminosäuren umfassende sogenannte RFX-DNA-Bindedomäne, deren Konsensus-Sequenz keine Homologie zu anderen bekannten DNA-Bindedomänen aufweist. Daneben enthalten die meisten Mitglieder der Proteinfamilie eine putative Dimerisationsdomäne, die weniger stark konserviert ist. Durch Homologievergleiche der RFX-Proteine konnten weitere konservierte Proteinregionen (A, B und C) lokalisiert werden, deren Funktion unbekannt ist. Außerdem enthalten einige Mitglieder der Familie Bereiche, in denen bestimmte Aminosäuren besonders häufig vorkommen. Das CPCR1-Protein besitzt die konservierten Regionen B und C, sowie im N-terminalen Bereich eine Prolin- und Glutamin-reiche und im C-terminalen Teil eine Asparaginsäure- und Glutaminsäure-reiche Region (Abbildung 3).

Die Deletion der C-terminalen 1,3 kb des CPCR1-Proteins in dem Vektor pGC1 führt zu dem Plasmid pGC1 Δ DIM1. Bei dieser Deletionsmutante sind die Nukleotide 1213 bis 2372 gemäß Tabelle 6 deletiert. Dies bedeutet, daß das mit dem Plasmid pGC1 Δ DIM1 exprimierbare Protein die Aminosäuresequenz von Position 1 bis Position 402 gemäß Tabelle 6 hat.

Hefezellen des Stammes BSII (siehe Tabelle 2) mit dem Plasmid pGC1 Δ DIM1 sind im Gegensatz zu solchen mit dem Ausgangsplasmid pGC1 nicht mehr in der Lage auf HIS-Selektionsplatten mit 55 mM 3-Aminotriazol zu wachsen. Dies deutet darauf hin, daß die DNA-Bindeaktivität des CPCR1-Proteins von der Dimerisationsdomäne abhängt. Möglicherweise bindet der Transkriptionsfaktor in diesem *in vivo*-Experiment nur als Homodimer an die DNA-Zielsequenz und führt somit zur Expression des HIS3-Reportergens.

Beispiel 3:

In vivo-Analysen zur Spezifität der DNA-Protein-Interaktion des CPCR1-Proteins

Transkriptionsfaktoren sind DNA-Bindeproteine, die in den meisten Fällen durch eine spezifische Interaktion mit DNA-Sequenzen eines Promotors an der transkriptionellen Regulation der Genexpression beteiligt sind. Zur exakten Bestimmung der Nukleotide, die für die spezifische Interaktion des CPCR1-Proteins mit Bereichen des *pcbAB/pcbC*-Promotors notwendig sind, wurden Mutationen in die präsumptive Bindestelle eingebracht. Es wurden hierzu Oligonukleotide synthetisiert (BSIIIm1-3), deren Sequenz jeweils an einer Stelle über einen Bereich von 2 bis 5 Nukleotiden von der Ausgangssequenz (BSII) abweicht. In der Tabelle 2 sind die Sequenzen der mutierten Oligonukleotide angegeben. Die dargestellten Sequenzen wurden in dreifacher Kopie synthetisiert und wie in Beispiel 1 beschrieben in das Hefegenom integriert. Die Reporterogenaktivität des *lacZ*-Gens bei Expression des CPCR1-Proteins in den resultierenden Hefestämmen ist in der Tabelle 2 angegeben. Die Substitution der Nukleotide CCAAT zu GGTTA in dem Oligonukleotid BSIIIm1 führt zu einem Verlust der *lacZ*-Aktivität des entsprechenden Hefestammes. Die Nukleotide CAA bilden möglicherweise die rechte Hälfte einer palindromischen Erkennungssequenz für den Faktor CPCR1. Bei einer Substitution der entsprechenden Nukleotide (TTG zu AAC) in der putativen linken Hälfte des Palindroms ist ebenfalls ein Verlust der *lacZ*-Aktivität festzustellen (siehe BSIIIm3). Diese Ergebnisse sprechen dafür, daß die Erkennungssequenz des Transkriptionsfaktors CPCR1 eine palindromische Struktur aufweist. Eine Substitution der Nukleotide GT in dem Oligonukleotid BSIIIm2 zerstört das Palindrom nicht vollständig und bewirkt auch keinen Verlust der *lacZ*-Aktivität.

Beispiel 4:

Nachweis des *cpcR1*-Gens in *Acremonium chrysogenum* durch DNA-Hybridisierungen

In der Figur 4 ist eine Southern-Hybridisierung unter stringenten Bedingungen mit genomischer DNA des *Acremonium chrysogenum* Stammes A3/2 dargestellt. Die DNA wurde mit den Restriktionsenzymen *EcoRI*, *BamHI* bzw. *HindIII* geschnitten und gelelektrophoretisch aufgetrennt. Als Sonde diente ein 2,4 kb-Fragment der *cpcR1*-cDNA, welches mit 32 Phosphor radioaktiv markiert wurde. Bei dem 2,4 kb großen Hybridisierungsfragment handelt es sich um ein *PstI*-Fragment, welches innerhalb des rekombinanten Plasmids pGC1 folgende Position einnimmt: das Fragment erstreckt sich von der *PstI* Erkennungsstelle an der Nukleotidposition 129 innerhalb des offenen Leserahmens des *cpcRI*-Gens bis zur *PstI*-Erkennungsstelle innerhalb der multiplen Klonierungsstelle des rekombinanten Plasmids. Die multiple Klonierungsstelle stammt von dem Ausgangsvektor pGAD424. Somit sind in dem Hybridisierungsfragment alle Genbereiche enthalten, mit Ausnahme der ersten 129 Nukleotide. Da in der genomischen DNA jeweils nur 1 Fragment markiert wird (7 bis ca. 15 kb), liegt das *cpcR1*-Gen wahrscheinlich in Einzelkopie im Genom von *A. chrysogenum* vor.

Hybridisierungspuffer: 50 % Formamid; 5x SSPE; 0,2 % SDS; 100 µg/ml

Heringssperma-DNA; Hybridisierungsbedingungen: 1 h Vorhybridisierung bei 37°C; 12 h Hybridisierung mit radioaktiv (32 P)-markierter Sonde; Waschpuffer: 5x SSPE, 0,2 % SDS; Waschbedingungen: 10 min bei 60°C.

Beispiel 5:

Nachweis des *cpcR1*-Gentranskripts in *Acremonium chrysogenum* durch RNA-Hybridisierungen

Die Expression des *cpcR1*-Gens in 3 verschiedenen *Acremonium chrysogenum* Stämmen wurde mittels einer Northern-Hybridisierung untersucht (Figur 5). Das als

Sonde eingesetzte radioaktiv-markierte Fragment des *cpcR1*-Gens (s. o.) markiert in der Gesamt-RNA von *A. chrysogenum* ein RNA-Molekül von etwa 3,5 kb. Die Unterschiede in der Intensität der Hybridisierung zwischen den 3 Pilzstämmen (*A. c.* ATCC 14553, A3/2 und Ac1) deuten auf eine unterschiedlich starke Expression des *cpcR1*-Gens hin. Eine höhere Expression des Transkriptionsfaktors CPCR1 in den beiden Semi-Produzentenstämmen A3/2 und Ac1 korreliert mit der erhöhten Cephalosporin-Syntheseleistung dieser Stämme. Dieses Ergebnis weist auf die mögliche regulatorische Wirkung des CPCR1-Proteins für die Cephalosporin C-Biosynthese hin. Hybridisierungspuffer: 50 % Formamid; 5x SSPE; 0,2 % SDS; 200 µg/ml Heringssperma-DNA; Hybridisierungsbedingungen: 2 h Vorhybridisierung bei 37°C; 12 h Hybridisierung mit radioaktiv (³²P)-markierter Sonde; Waschpuffer: 5x SSPE, 0,2 % SDS; Waschbedingungen: 10 min bei 60°C.

Beispiel 6:

Konstruktion von Expressionsvektoren für *Acremonium chrysogenum*

Im folgenden werden die Expressionsvektoren pIR11, pIR12, pIR13, pIR21, pIR22, pIR23, pIR31, pIR32 und pIR33 beschrieben. Die Konstruktion der Vektoren wurde nach Standardmethoden der molekularen Genetik vorgenommen (Sambrook et al. 1989). Die Nukleotidsequenz der Vektoren ist aus Tabelle 10 abzuleiten.

Allgemeine Beschreibung der Vektoren:

1. Alle Vektoren basieren auf dem bakteriellen Plasmid PCRScript+AmpSK(+) (Stratagene, Heidelberg) und erhalten für die Vermehrung in *E. coli* einen Replikationsursprung, welcher in *E. coli* wirksam ist. Außerdem tragen sie ein Ampicillin-Resistenzgen, welches für eine β-Lactamase kodiert. Die Selektion in *E. coli* wird durch das Resistenzgen ermöglicht. In Tabelle 10 wird der PCRScript+AmpSK(+)-Anteil als "bakterieller Anteil" bezeichnet.

2. Alle Vektoren enthalten den Promotorbereich des pcbC-Gens (Samson et al. 1985, Menne et al. 1994).
3. Alle Vektoren enthalten die Transkriptions-Terminationssignale des pcbC-Gens (Samson et al. 1985)
4. Die Vektoren enthalten variable Klonierungsinserionsschnittstellen, die durch die Linker der Tabelle 11 bestimmt werden.

Die drei unterschiedlichen Klonierungsinserionsstellen unterscheiden sich lediglich durch einzelne Nukleotide (unterstrichen in der Tabelle 11). Dadurch ist eine

Translationsfusion von zwei Gensequenzen in den drei möglichen Leserahmen eines Protein-kodierenden Gens möglich.

5. Die Vektoren pIR21, pIR22, pIR23 und pIR31, pIR32, pIR33 enthalten Sekretions-Signalsequenzen für Translationsfusionen.

Um eine effiziente Expression und eine möglichst einfache Reinigung des heterologen Genprodukts zu ermöglichen, wurden unmittelbar vor der Klonierungsinserionsstelle sogenannte Sekretions-Signalsequenzen eingefügt, die einen Transport des exprimierten Proteins in das endoplasmatische Reticulum und schließlich eine Sekretion aus den Zellen bewirken. Das Genprodukt sollte somit in den Kulturüberständen der entsprechenden Pilztransformanten vorliegen. Es wurden zwei verschiedene Export-Signalsequenzen verwendet:

- a) Zum einen handelt es sich um eine Sequenz mit einer Länge von 42 bp, die Teil eines Gens ist, das für eine alkalische Protease aus dem Pilz *Fusarium sp.* kodiert. Das kodierte Signalpeptid hat eine Länge von 14 Aminosäuren und wurde bereits in *Saccharomyces cerevisiae* und in *A. chrysogenum* auf seine Funktionalität

getestet, jedoch nicht in Verbindung mit einem heterologen Gen (Kitano et al. 1992). Die Signalsequenz ist in den Vektoren pIR21, pIR22 und pIR23 enthalten.

b) Die zweite verwendete Signalsequenz stammt aus *A. chrysogenum* und ist ebenfalls Teil eines Gens, das für eine alkalische Protease kodiert. Diese Sequenz hat eine Länge von 60 bp, wobei das resultierende Signalpeptid, dessen Funktionalität gleichfalls in *S. cerevisiae* und *A. chrysogenum* verifiziert wurde, eine Länge von 20 Aminosäuren hat (Isogai et al. 1991). Die Signalsequenz aus *A. chrysogenum* ist in den Vektoren pIR31, pIR32, pIR33 enthalten.

Zusammenfassend werden die Übergangssequenzen der neun Expressionsvektoren in der Tabelle 12 angegeben.

Beispiel 7:

Expression des CPCR1 Proteins in *A. chrysogenum*

Die im Beispiel 6 beschriebenen Vektoren können benutzt werden, um das CPCR1 Protein in *A. chrysogenum* zu exprimieren.

Zu diesem Zweck kann das cpcR1-Gen aus dem Plasmid pGC1 durch PCR-Amplifikation (Sambrook et al. 1989) synthetisiert werden. Bei dieser Amplifikation werden die folgenden zwei Primer eingesetzt (in Klammern sind die Positionen der Oligonukleotide innerhalb der Nukleotidsequenz gemäß Tabelle 6 genannt).

1. 5'CCT GCT ATG TAC GGA CAA GGG3' (10-30) (SEQ ID NO.: 54)
2. 5'CC TCG TCA TGC AGG AGC CGC CC3' (2351-2372) (SEQ ID NO.: 55)

Durch die Amplifikation erhält man ein 2362 bp großes DNA-Fragment, das durch DNA-Sequenzierung auf seine Korrektheit hin überprüft wird. Dieses Fragment wird in die NruI-Restriktionsstelle des Plasmids pIR11 kloniert. Durch das Enzym NruI wird ein sogenanntes "stumpfes Ende" hergestellt, mit dem die Klonierung des PCR-

Fragments ohne vorausgegangene Restriktion möglich wird. Durch die Ligation des Vektors und des PCR-Fragments entsteht ein rekombinantes Plasmid mit der Bezeichnung pIRS11, das anschließend in *E. coli* kloniert und vermehrt werden kann.

Durch anschließende DNA-Kotransformationsexperimente (Menne et al. 1994) wird das Plasmid in *A. chrysogenum* eingebracht. In der Regel erfolgt eine ektopische Integration des Plasmids in mehreren Kopien in das pilzliche Genom. Es können also pilzliche Transformanten mit mehreren Kopien des *cpcRI*-Gens unter der Kontrolle des *pcbC*-Promotors erhalten werden. Die anschließende molekulare Analyse der Pilztransformanten erlaubt die Selektion geeigneter Stämme. Diese sollten das *CPCR*¹-Protein übersynthetisieren und somit eine erhöhte Expression der Cephalosporin C-Biosynthesegene aufgrund der *CPCR*¹-Wirkung zur Folge haben.

Beispiel 8:

Charakterisierung der Dimerisationsdomäne des *CPCR*1-Proteins

Das *CPCR*1-Protein besitzt vermutlich mindestens zwei wichtige funktionelle Domänen, die DNA-Bindedomäne und die Dimerisationsdomäne. Die Charakterisierung der potentiellen Dimerisationsdomäne wird im folgenden durch *in vivo* Experimente mit Derivaten des *CPCR*1-Proteins beschrieben. Die Analysen wurden mit geeigneten Hefestämmen in Form von ONE-HYBRID und TWO-HYBRID Versuchen durchgeführt.

A. ONE-HYBRID Experimente zur *in vivo* DNA-Bindeaktivität von Wildtyp und mutiertem *CPCR*1-Protein.

Mit dem ONE-HYBRID-System kann die DNA-Protein-Interaktion *in vivo* untersucht werden. Dabei besteht die Möglichkeit, die DNA-Bindesequenz zu variieren und so eine Aussage über Nukleotide zu erhalten, welche für die Protein-Bindung essentiell

sind. Andererseits kann auch der Protein-Anteil des analysierten Komplexes verändert werden. In beiden Fällen wird die Stärke der Interaktionen durch eine Messung der Reporterogenaktivität ermittelt.

Die Reporterogenaktivität von Hefetransformanten im ONE-HYBRID-System bei Expression des *cpcR1*-Gens und dessen Deletionskonstrukten sind in der Figur 6 dargestellt. Das Plasmid pGC1 kodiert für ein Fusionsprotein aus der GAL4-AD und dem CPCR1-Protein von *A. chrysogenum*. Die Transformation dieses Plasmids in den Reporterengestamm mit der Bindestelle II aus dem *pcbC*-Promotor resultiert in einer hohen β -Galaktosidase-Aktivität (siehe Figur 6).

Während der Interaktion des CPCR1-Proteins mit der Bindestelle erfolgt eine Aktivierung der Transkription des Reportergens durch die fusionierte GAL4-

Domäne. Die Expression des *cpcR1*-Gens ohne die aktivierende Domäne (CPCR1) führt wie in der Figur 6 zu erkennen nicht zu einer erhöhten Reporterogenaktivität. In der Abbildung sind außerdem die Auswirkungen von Deletionen innerhalb des *cpcR1*-Gens dargestellt. Das Protein AD-DIM1 besitzt nur die N-terminale Hälfte des CPCR1-Proteins. Dadurch fehlt die Dimerisationsdomäne vollständig, während die DNA-Bindedomäne erhalten bleibt. Hefetransformanten, in denen dieses Derivat des CPCR1-Proteins gebildet wird, zeigen keine Reporterogenaktivität, was auf eine fehlende DNA-Bindung hinweist. Die Auswirkungen einer kleineren Deletion sind bei dem Protein AD-DIM2 erkennbar. Hier beginnt die Deletion in der Mitte der Dimerisationsdomäne, so daß die N-terminale Hälfte der Domäne erhalten bleibt. Die β -Galaktosidase-Aktivität eines Proteinextraktes von Hefetransformanten mit diesem Konstrukt liegt bei 32 U/mg Protein. Dies entspricht etwas mehr als 5 % der Aktivität des Wildtyp-Proteins. Diese ONE-HYBRID Versuche zeigen, daß Deletionen des CPCR1-Proteins zu einem unterschiedlich starken Verlust der analysierten DNA-Bindeaktivität führen.

Der Einsatz von Derivaten des CPCR1-Proteins im ONE-HYBRID-System erlaubt die Analyse funktioneller Proteindomänen. Die Dimerisationsdomäne des CPCR1-

Proteins scheint eine entscheidende Rolle für die DNA-Bindeaktivität des Proteins zu besitzen, denn eine vollständige oder unvollständige Deletion dieser Domäne führt zu einer starken Abnahme der Reportergenaktivität.

B. TWO-HYBRID Experimente zur Bestimmung von Protein-Protein-Interaktionen zwischen dem CPCR1-Protein und dessen Derivaten

Das TWO-HYBRID-System ermöglicht den direkten Nachweis von Protein-Protein-Interaktionen (Fields und Song 1989). Es wird, wie das ONE-HYBRID-System, in der Hefe *S. cerevisiae* durchgeführt. Die beiden auf eine mögliche Interaktion zu testenden Proteine müssen jeweils als Fusionsproteine mit zwei verschiedenen Domänen des GAL4-Proteins der Hefe in einen geeigneten Hefestamm eingebracht werden. Die DNA-Bindedomäne des GAL4-Proteins (BD) wird mit dem einem

Interaktionspartner fusioniert, die aktivierenden Domäne des GAL4-Proteins (AD) mit dem anderen Protein. Interagieren die beiden Proteine miteinander, gelangen die beiden GAL4-Domänen in räumliche Nähe zueinander und es kommt zur Expression von Reportergenen in den entsprechenden Hefetransformanten. Die im vorherigen Abschnitt verwendeten Derivate des CPCR1-Proteins wurden auch im TWO-HYBRID-System eingesetzt. Dazu wurden das CPCR1-Protein und die Proteine DIM1 und DIM2 (siehe Figur 6) mit der DNA-Bindedomäne (BD) des GAL4-Proteins fusioniert. Zusammen mit den bereits vorhandenen Konstrukten, die eine Fusion mit der aktivierenden Domäne (AD) darstellen, können diese Proteine auf Interaktionen untereinander getestet werden.

Es wurden sämtliche Kombinationen der bereits beschriebenen Proteine untersucht. Zu diesem Zweck wurden zunächst drei Stämme erzeugt, die jeweils die Fusionsproteine BD-CPCR1, BD-DIM1 und BD-DIM2 exprimierten. Anschließend wurden in diese Stämme entweder eines der drei entsprechenden Konstrukte mit den AD-Fusionen oder eines von zwei Kontrollplasmiden transformiert. Bei den Kontrollplasmiden handelt es sich um das Plasmid pGAD424, welches nur für die AD kodiert und um das Plasmid pC1, welches das *cpcR1*-Gen ohne eine fusionierte

GAL4-Domäne enthält. In der Figur 7A befindet sich eine schematische Darstellung der fünf Hefestämme mit den verschiedenen Proteinen.

In dem verwendeten Hefestamm ist die Bindestelle des GAL4-Proteins stromaufwärts des *HIS3*-Gens integriert, so daß eine Interaktion der untersuchten Proteine zur Histidin-Prototrophie führt. Entsprechend wurde die Analyse der Protein-Protein-Interaktion in den Hefetransformanten auf Selektionsplatten durchgeführt (siehe Figur 7). Vom Wachstum der Hefezellen kann auf die Stärke der Interaktion geschlossen werden. Ein ganz schwaches Wachstum entlang der Impfstriche geht auf die Hintergrundaktivität des *HIS3*-Gens in den Hefezellen zurück.

Es wird deutlich, daß die Kombination von CPCR1 mit unverändertem CPCR1 zur stärksten Wechselwirkung führt (Figur 7B). Die Deletion im Protein DIM2 erlaubt noch eine Interaktion mit sich selbst, wenngleich die Affinität der Interaktionspartner

sehr gering ist (Figur 7D). Das Protein DIM1 mit der größeren Deletion kann dagegen nicht mehr mit sich selbst interagieren (Figur 7C).

In der Abbildung 15B und D ist auch zu erkennen, daß das vollständige CPCR1-Protein mit beiden Derivaten DIM1 und DIM2 interagieren kann. Allerdings ist die Interaktion mit DIM2 wesentlich stärker als mit DIM1. Die Kontrollproteine, AD und CPCR1 ohne die Fusion, führen in keinem Fall zu einem Wachstum der Hefetransformanten.

Durch TWO-HYBRID Versuche konnte gezeigt werden, daß das CPCR1-Protein mit sich selbst interagiert. Der Bereich der Dimerisationsdomäne ist für eine starke Interaktion notwendig, wenngleich eine schwache Interaktion wahrscheinlich auch über die N-terminale Hälfte des Proteins vermittelt werden kann.

Beispiel 9:

Nachweis einer spezifischen Bindung des CPCR1-Proteins an eine palindromische Sequenz des *pcbC*-Promotors

Bereits beim Durchsuchen der cDNA-Bank nach möglichen positiven Hefetransformanten wurde neben der Wildtyp-Bindestelle BSII aus dem *pcbC*-Promotor auch die mutierte Bindestelle BSII_{m1} als Spezifitätskontrolle eingesetzt. Dadurch konnte überprüft werden, ob der isolierte putative Transkriptionsfaktor von *A. chrysogenum* mit der Bindestelle in den Hefezellen sequenzspezifisch interagiert ist. Weitere Mutationen und Gelretentionsanalysen erlauben eine genauere Bestimmung der Bindestelle des CPCR1-Proteins.

A. In vivo Analysen des CPCR1-Proteins mit mutierten Bindestellen unter Verwendung des ONE-HYBRID-Systems

Das 24 Nukleotide umfassende DNA-Fragment BSII aus dem *pcbC*-Promotor, welches als potentielle Bindestelle eines Transkriptionsfaktors für den Einsatz im ONE-HYBRID-System ausgewählt wurde, enthält mit der Basenabfolge 5'-CCAAT-3'

eine putative CCAAT-Box. Einige der Nukleotidpositionen in der CCAAT-Box sind zusätzlich Teil einer palindromischen Struktur. In der Figur 2 ist diese Sequenz wiedergegeben (BSII), wobei Nukleotide unterstrichen sind, die Teil des Palindroms sind. Mit Hilfe von Hefe-Reportergenstämmen wurde untersucht, wie sich Mutationen in der CCAAT-Box und anderen Bereichen des Palindroms auf die *in vivo* DNA-Protein-Interaktion im ONE-HYBRID-System auswirken. Dazu wurden die in Figur 2 dargestellten mutierten Bindestellen BSII_{m2} und BSII_{m3} in je dreifacher Kopie vor die Reportergene *HIS3* und *lacZ* integriert, so daß zusammen mit BSII_{m1} für jedes Reportergen insgesamt drei Reportergenstämmen mit mutierten Bindestellen zur Verfügung standen. Bei dem Stamm mit der Bindestelle BSII_{m1} handelt es sich um den bereits beschriebenen Reportergenstamm, der zur Selektion des cDNA-Klons eingesetzt wurde (siehe Beispiel 1). Nach der Transformation des

Expressionsplasmids mit der cDNA des *cpcR1*-Gens in die genannten Reportergenstämme konnte die Aktivität der Reportergene analysiert werden.

Die Aktivität des *HIS3*-Gens in den Transformanten wurde durch einen Wuchstest auf Selektionsplatten, die Expression des *lacZ*-Gens durch eine Messung der Enzymaktivität bestimmt. Proteinextrakte aus Hefestämmen mit der Wildtyp-Sequenz BSII zeigen eine Aktivität der β -Galaktosidase von 603 U/mg Proteinextrakt. In dem Stamm mit der Sequenz BSII_{m1} ist die gesamte CCAAT-Box mutiert und damit auch die rechte Hälfte des Palindroms. In Transformanten dieses Stammes ist die Aktivität der β -Galaktosidase mit neun U/mg Proteinextrakt sehr gering. Auch eine Mutation der linken Hälfte des Palindroms (BSII_{m3}, 5,7 U/mg) führt zu einem Verlust der Reportergenexpression, was auf eine gestörte DNA-Protein-Interaktion hinweist. In der Sequenz BSII_{m2} werden innerhalb der CCAAT-Box nur zwei Nukleotide ausgetauscht. Die resultierende Aktivität der β -Galaktosidase beträgt mit 140 U/mg Proteinextrakt weniger als ein Viertel der Wildtyp-Aktivität.

Reportergenanalysen mit mutierten Bindestellen im ONE-HYBRID-System zeigten, daß das CPC_{R1}-Protein spezifisch an eine palindromische Sequenz aus dem *pcbC*-

Promotor bindet. Die rechte Hälfte des Palindroms fällt mit einer CCAAT-Box zusammen. Beide Hälften des Palindroms sind für eine Bindung essentiell.

B. In vitro Gelretentionsanalysen mit Proteinextrakten aus *A. chrysogenum* und rekombinanten Hefestämmen

Die Interaktion zwischen einem Protein und einem bekannten DNA-Fragment kann auch *in vitro* mittels der Gelretentionsanalyse untersucht werden, bei der radioaktiv markierte DNA, welche die Protein-Bindestelle enthält, mit Proteinextrakten inkubiert wird. Anschließend werden DNA-Protein-Komplexe gelelektrophoretisch von freien DNA-Molekülen getrennt und durch Autoradiographie nachgewiesen.

Ein doppelsträngiges Oligonukleotid, welches dreimal die Sequenz der Bindestelle BSII (siehe Figur 2) enthält, wurde durch den Einbau radioaktiver Nukleotide markiert und mit verschiedenen Proteinextrakten inkubiert. Protein-Rohextrakte wurden aus

A. chrysogenum und rekombinanten Hefestämmen gewonnen und durch „Fast-Protein-Liquid“-Chromatographie (FPLC) unter Verwendung einer Heparinsäule aufgereinigt und fraktioniert.

Mit Proteinextrakt aus *Acremonium* und aus einem Hefestamm, der das CPCR1-Protein bildet, konnte eine spezifische DNA-Protein-Interaktion mit der Bindestelle II nachgewiesen werden. Die Spezifität der Bindung wurde durch Kompetitionsanalysen untersucht. Durch die Zugabe von nicht radioaktiv markierten DNA-Molekülen zum Reaktionsansatz kann getestet werden, ob das Protein auch an diese DNA-Moleküle bindet und damit im verminderten Umfang den spezifischen DNA-Protein-Komplex bildet. Eine Konkurrenz mit der Bindestelle II führt zu einer starken Abnahme des Komplexes. Bei Zugabe von Oligonukleotiden mit einer mutierten Sequenz (BSII_{m3}) kann dies nicht beobachtet werden. Dies deutet auf eine spezifische Interaktion zwischen dem CPCR1-Protein und der Bindestelle II hin.

Des weiteren wurden Gelretentionsanalysen mit 2 weiteren Hefestämmen vorgenommen. Der eine Hefestamm exprimiert das CPCR1-Protein als

Fusionsprotein mit der transkriptionsaktivierenden Domäne des GAL4-Proteins der Hefe (AD-CPCR1). Der andere Stamm wurde als Kontrolle verwendet und exprimiert nur die Domäne des GAL4-Proteins (AD). Dieser Kontrollproteinextrakt führt nicht zur Bildung des spezifischen DNA-Protein-Komplexes, der mit dem Protein AD-CPCR1 beobachtet werden kann. Durch diesen Vergleich der Hefeproteinextrakte konnte gezeigt werden, daß der Proteinextrakt AD-CPCR1 aus dem rekombinanten Hefestamm für *in vitro* Bindungsstudien geeignet ist.

Durch Gelretentionsanalysen konnte eine spezifische Interaktion von Proteinextrakten aus *A. chrysogenum* und rekombinanten Hefestämmen mit der

Bindestelle II nachgewiesen werden. Damit bestätigt sich die Spezifität der Bindung des CPC1-Proteins an die palindromische Sequenz aus dem *pcbC*-Promotor auch *in vitro*.

Beispiel 10:

Genomische DNA von *A. chrysogenum* enthaltend das CPC1-Gen

Es wurde ein genomisches DNA-Fragment aus *A. chrysogenum* isoliert, welches das CPC1-Gen enthält. Die Sequenz ist in Tabelle 13 gezeigt (SEQ ID NO.: 57). Der offene Leserahmen beginnt bei Position 535 und endet bei Position 3130. Es sind 2 Introns enthalten: Intron 1 von Position 1215 bis Position 1267, Intron 2 von Position 1568 bis Position 1619.

Figurlegenden:

Fig. 1: Schema des ONE-HYBRID-SYSTEMS. Ein Hefe-Reportergenstamm wird mit einer Expressionsgenbank von *A. chrysogenum* transformiert. Bei einer DNA-Protein-Interaktion zwischen dem heterologen Anteil des Fusionsproteins und dem analysierten DNA-Element (E) kommt es zur Expression des Reportergens (HIS3). Dieses Prinzip der Selektion auf Histidin-Auxotrophie erlaubt die direkte Identifizierung von Transformanten, deren Expressionsplasmid für einen möglichen

trans-Faktor kodiert. Abkürzungen: AD: transkriptionsaktivierende GAL4-Domäne, BP: DNA-Bindeprotein, E: analysiertes DNA-Element, HIS3: Histidin-Biosynthesegen, P: Promotor

Fig. 2: Schematische Darstellung des intergenischen Promotorbereiches der *pcbAB/pcbC*-Gene von *Acremonium chrysogenum*. Der Promotorbereich ist schraffiert, die Translationsstartpunkte der beiden divergenten Gene sind durch Pfeile markiert. Dargestellt ist die Sequenz (Position -447 bis -423) des Promotors, welche in dem ONE-HYBRID-SYSTEM zur Selektion spezifischer

Transkriptionsfaktoren eingesetzt wurde, sowie Variationen davon. Ebenfalls dargestellt wird die quantitative Analyse der Aktivität der β -Galaktosidase von entsprechenden Reportergenstämmen.

Fig. 3: Schematische Darstellung des CPCR1-Proteins. Die DNA-Bindedomäne (DBD) und die Dimerisationsdomäne (DIM) sind durch einen fettgedruckten Rahmen gekennzeichnet. Außerdem sind die innerhalb der RFX-Familie konservierten Regionen B und C eingezeichnet. Proteinregionen, die einen erhöhten Anteil bestimmter Aminosäuren enthalten sind mit PQ (Prolin und Glutamin) bzw. mit DE (Asparaginsäure und Glutaminsäure) markiert.

Fig. 4: Southern-Hybridisierung von Gesamt-DNA des *A. chrysogenum*-Stammes Ac1. Die Gesamt-DNA wurde mit Restriktionsenzymen wie angegeben geschnitten und gelelektrophoretisch aufgetrennt. Als Sonde wurde ein radioaktiv-markiertes *Pst*I-Fragment des *cpcR1*-Gens mit einer Größe von 2,4 kb eingesetzt. Als Positivkontrolle für die Hybridisierung diente das Plasmid pGC1, welches die *cpcR1*-cDNA enthält.

Fig. 5: Northern-Hybridisierung mit Gesamt-RNA der *A. chrysogenum*-Stämme ATCC 14553 (Spur1), A3/2 (Spur2) und Ac1 (Spur3). Als Sonde wurde ein radioaktiv-markiertes *Pst*I-Fragment des *cpcR1*-Gens mit einer Größe von 2,4 kb eingesetzt.

Fig. 6: Schematische Darstellung der im ONE-HYBRID-System eingesetzten Proteine und der β -Galaktosidase-Aktivität der entsprechenden Hefetransformanten. Die β -Galaktosidase-Aktivität wurde aus mindestens drei Experimenten ermittelt. Wichtige Abkürzungen: AD: transkriptionsaktivierende Domäne des GAL4-Proteins, DBD: DNA-Bindedomäne, DIM: Dimerisationsdomäne, DIM1/2: Deletionskonstrukte des *cpcR1*-Gens, U β -Gal: U β -Galaktosidase pro mg Protein. Alle weiteren Bezeichnungen entsprechen den Anmerkungen zu Figur 3.

Fig. 7: Reporteranalyse von TWO-HYBRID Experimenten mit dem CPC1-Protein und deletierten Derivaten des CPC1-Proteins. A) Schematische Darstellung der Hefestämme mit den jeweiligen Proteinen. B), C) und D) Photographien von Histidin-Selektionsplatten. Interaktionen der untersuchten Proteine führen in Hefetransformanten zu einer Expression des *HIS3*-Gens, die den Hefezellen das Wachstum ermöglicht. Die Selektionsplatten enthalten kein Histidin, Leucin und Urazil, aber 6 mM 3-Aminotriazol. Abkürzungen: AD: aktivierende Domäne, BD: DNA-Bindedomäne. Schematische Darstellung der Proteine siehe Figur 6.

Literatur

Cheng H, Kinsey JA (1995) Purification of a heteromeric CCAAT binding protein from *Neurospora crassa*. Mol Gen Genet 249: 301-308

Chiang T, Marzluff GA (1995) Binding affinity and functional significance of NIT2 and NIT4 binding sites in the promoter of the highly regulated *nit-3* gene, which encodes nitrate reductase in *Neurospora crassa*. J Bacteriol 177: 6093-6099

Clontech (1995) Matchmaker One-Hybrid System protocol. Catalog # K1603-1

Emery P, Durand B, Mach B, Reith W (1996) RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. Nucl Acids Res 24: 803-807

Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature 340: 245-246

Fu YH, Paietta JV, Mannix DG, Marzluf GA (1989) Cys-3, the positive-acting sulfur regulatory gene of *Neurospora crassa*, encodes a protein with a putative leucine zipper DNA-binding element. Mol Cell Biol 9:1120-1127

Isogai T, Fukagawa M, Kojo H., Kohsaka M, Aoki H, Imanaka H (1991) Cloning and nucleotide sequences of the complementary and genomic DNAs for the alkaline protease from *Acremonium chrysogenum* Agric. Biol. Chem. 55:471-477

Kang S, Metzenberg RL (1990) Molecular ananlysis of *nuc-1+*, a gene controlling phosphorus acquisition in *Neurospora crassa*. Mol Cell Biol 10: 5839-5848

Kitano K, Morita S, Kuriyama M, Mejima K (1992) Alkaline protease gene from *Fusarium*. European Patent No. 0519229A2

Kudla B, Caddick MX, Langdon T, Martinez-Rossi NM, Bennett CF, Sibley S, Davis RW, Arst HN Jr (1990) The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. EMBO J 9: 1355-1364

Marzluff GA (1993) Regulation of sulfur and nitrogen metabolism in filamentous fungi. Annu Rev Microbiol 47: 31-55

Menne S, Walz M, Kück U (1994) Expression studies with the bidirectional *pcbAB*-*pcbC* promoter region from *Acremonium chrysogenum* using reporter gene fusions. Appl. Microbiol. Biotechnol. 42: 57-66

Nosek J, Radzio R, Kück U (1997) Produktion von β -Lactamantibiotika durch Mikroorganismen. Chemie in unserer Zeit 31: 172-182

Nussinov R (1992) The eukaryotic CCAAT and TATA boxes, DNA spacer flexibility and looping. J Theor Biol 155: 243-270

Pabo CO, Sauer RT (1992) Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem 61: 1053-1095

Payne GA, Nystrem GJ, Bhatnagar D, Cleveland TE, Woloshuk CP (1993) Cloning of the *affl-2* gene involved in aflatoxin biosynthesis from *Aspergillus flavus*. Appl Environ Microbiol 59: 156-162

Radzio R (1996) Heterologe Genexpression in dem Cephalosporin C-produzierendem Hyphenpilz *Acremonium chrysogenum*. Dissertation, Ruhr-Universität Bochum

Radzio R, Kück U (1997) Efficient synthesis of the blood-coagulation inhibitor hirudin in the filamentous fungus *Acremonium chrysogenum*. Appl. Microbiol Biotechnol (199) 48: 58-65

Raymondjean M, Cereghini S, Yaniv M (1988) Several distinct „CCAAT“ box binding proteins coexist in eukaryotic cells. Proc Natl Acad Sci 85: 757-761

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning - A laboratory manual. Cold Spring Harbour Laboratory Press, New York

Samson SM, Belagaje R, Blankenship DT, Chapman JL, Perry D, Skatrud PL, VanFrank RM, Abraham EP, Baldwin JE, Queener SW, Ingolia TD (1985) Isolation, sequence determination and expression in *Escherichia coli* of the isopenicillin N synthetase gene from *Cephalosporium acremonium*. Nature 318: 191-194

Schjerling P, Holmberg S (1996) Comparative amino acid sequence analysis of the C₆ zinc cluster family of transcriptional regulators. Nucl Acids Res 24: 4599-4607

Then Bergh K, Litzka O, Brakhage AA (1996) Identification of a major *cis*-acting DNA element controlling the bidirectionally transcribed penicillin biosynthesis genes *acvA* (*pcbAB*) and *ipnA* (*pcbC*) of *Aspergillus nidulans*. J Bacteriol 178: 3908-3916

Wang MM, Reed RR (1993) Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature* 364: 121-126

Tabellen

Tab. 1: Exemplarische Auswahl von Transkriptionsfaktoren bei Hyphenpilzen.

Abkürzungen: bZIP: basic Leucin-Zipper, bHLH: basic Helix-Loop-Helix,

Transfaktor: transaktivierender Faktor

Organismus	Transfaktor	Klasse	Binde-Sequenz im Promotor	regulierte Gene	Referenz
<i>Aspergillus flavus</i>	AflR	C6-Zinkfinger	TTAGGCCTAA (SEQ ID NO.: 2)	Aflatoxin-Biosynthesegene	Payne et al. 1993
<i>Aspergillus nidulans</i>	AREA	C2/C2-Zinkfinger	GATA-Elemente	mehr als 100 Strukturgene des N ₂ -Metabolismus	Kudla et al. 1990
<i>Neurospora crassa</i>	CYS3	bZIP	ATGACGTCAT (SEQ ID NO.: 3)	<i>cys13</i> und <i>cys14</i>	Fu et al. 1989
	NUC-1	bHLH	CACGTG	<i>pho4+</i> , <i>pho5+</i> und <i>pho2+</i>	Kang und Metzenberg 1990
<i>Penicillium chrysogenum</i>	PacC	C2/C2-Zinkfinger	GCCAAG	<i>pcbAB/pcbC</i> , pH-Regulation	Haas und Marzluff 1995

Tab. 2: Monomere Teilsequenzen von Oligonukleotiden, die im ONE-HYBRID-SYSTEM eingesetzt wurden. Die Sequenz BSII entspricht dem Bereich von -447 bis -423 des *pcbAB/pcbC*-Promotors von *A. chrysogenum*. Die mutierten Oligonukleotide BSII_{m1-3} wurden zur Bestimmung der Bindesequenz des CPCR1-Proteins eingesetzt. Die von der Ausgangssequenz abweichenden Nukleotide sind in Fettdruck angegeben. Neben den Ergebnissen der Reportergenanalysen ist auch die Konsensus-Sequenz der Bindestellen des humanen RFX1-Proteins angegeben. Unterstrichen sind solche Nukleotide innerhalb der Sequenzen, die an der Bildung eines Palindroms beteiligt sein könnten.

Oligonukleotid	Sequenz	in vivo-Nachweis der lacZ-Reportergenaktivität
BSII		
(SEQ ID NO.: 4)	5' GTTGCCGGG <u>CCA</u> TCCCTGAGCTT3'	+
(SEQ ID NO.: 5)	CAACGGCCCGGTTAGGGACTCGAA	
BSII _{m1}		
(SEQ ID NO.: 6)	5' GTTGCCGGGGGTTACCCTGAGCTT3'	-
(SEQ ID NO.: 7)	CAACGGCCCCCAATGGGACTCGAA	
BSII _{m2}		
(SEQ ID NO.: 8)	5' GTTGCCGGGCGTATCCCTGAGCTT3'	+
(SEQ ID NO.: 9)	CAACGGCCCGCATAGGGACTCGAA	
BSII _{m3}		
(SEQ ID NO.: 10)	5' CAACCCGGGCCAATCCCTGAGCTT3'	-
(SEQ ID NO.: 11)	GTTGGGCCCGGTTAGGGACTCGAA	
RFX1-consensus		
(SEQ ID NO.: 56)	5' NNGTNRCCRGYAACNN3'	

Tab. 3: Sequenzen komplementärer Oligonukleotide mit dreifach Kopien der putativen Transkriptionsfaktor-Bindesequenz aus dem *pcbAB/pcbC*-Promotor. Oben ist die Wildtypsequenz angegeben, darunter Derivate mit Substitutionen der DNA-Bindestelle (unterstrichen). Die hybridisierten Oligonukleotide wurden in die Klonierungsstellen von Reportergenvektoren (pHISi und pLACi) integriert und mittels DNA-Transformation in das Hefegenom eingebracht.

Nam des Dimers	Sequenzen der hybridisierten Oligonukleotide
BSII	
(SEQ ID NO.: 12)	5'AATTC GTTGCCGGGGCCAAATCCCTGAGCTT GTTGCCGGGGCCAAATCCCTGAGCTT CCC 3'
(SEQ ID NO.: 13)	3' G CAACGGCCCCGGTTAGGGACTCGAA CAACGGCCCCGGTTAGGGACTCGAA GGG 5'
BSIIIm1	
(SEQ ID NO.: 14)	5'AATTC GTTGCCGGGGGGTTACCCCTGAGCTT GTTGCCGGGGGGTTACCCCTGAGCTT CCC 3'
(SEQ ID NO.: 15)	3' G CAACGGCCCCCCTCAATGGGACTCGAA CAACGGCCCCCCTCAATGGGACTCGAA GGG 5'
BSIIIm2	
(SEQ ID NO.: 16)	5'AATTC GTTGCCGGGGGGTATCCCTGAGCTT GTTGCCGGGGGGTATCCCTGAGCTT CCC 3'
(SEQ ID NO.: 17)	3' G CAACGGCCCCGCATAGGGACTCGAA CAACGGCCCCGCATAGGGACTCGAA GGG 5'
BSIIIm3	
(SEQ ID NO.: 18)	5'AATTC CAACCCGGGGCCAAATCCCTGAGCTT CAACCCGGGGCCAAATCCCTGAGCTT CCC 3'
(SEQ ID NO.: 19)	3' G GTTGGGGCCCCGGTTAGGGACTCGAA GTTGGGGCCCCGGTTAGGGACTCGAA GGG 5'

Tab. 4: Übersicht über die verwendeten Hefepasmide. Der Teil A der Tabelle enthält Hefepasmide mit Integrationen der in Tabelle 3 aufgeführten Oligonukleotide. Es handelt sich um integrative Hefepasmide. Der Teil B der Tabelle enthält Hefe-Expressionsplasmide, welche in die erzeugten Hefestämmen transformiert wurden und dort als autonom replizierende Plasmide vorliegen.

A	Plasmid	Ausgangsvektor	integrierte Oligonukleotide	Referenz
	pHISi			Clontech
	pHISi-BSII	pHISi	BSII	diese Arbeit
	pHISi-BSIIIm1	pHISi	BSIIIm1	diese Arbeit
	pHISi-BSIIIm2	pHISi	BSIIIm2	diese Arbeit
	pHISi-BSIIIm3	pHISi	BSIIIm3	diese Arbeit
	pLACZi			Clontech
	pLACZi-BSII	pLACZi	BSII	diese Arbeit
	pLACZi-BSIIIm1	pLACZi	BSIIIm1	diese Arbeit
	pLACZi-BSIIIm2	pLACZi	BSIIIm2	diese Arbeit
	pLACZi-BSIIIm3	pLACZi	BSIIIm3	diese Arbeit
B	Plasmid	Ausgangsvektor	Charakteristika	Referenz
	pGAD424		Klonierungsvektor für cDNA-Bank, GAL4-Aktivierungsdomäne	Clontech
	pGC1	pGAD424	cDNA-Insert: <i>cpcR1</i> -Gen in <i>EcoRI</i> / <i>BamHI</i> -Schnittstelle	diese Arbeit
	pGC1 Δ DIM1	pGC1	Deletion bp 1220 - 2380 des <i>cpcR1</i> -Gens	diese Arbeit

Tab. 5: Übersicht über die verwendeten Hefestämme. Die transgenen Hefestämme wurden durch die Transformation mit integrativen und/oder frei replizierenden Hefepiasmiden (siehe Tabelle 4) erzeugt.

<i>Stamm</i>	<i>im Genom integrierte Plasmide</i>	<i>autonom replizierende Plasmide</i>	<i>Referenz</i>
YM4271			Clontech
HISLACZ-BSII	pHISi-BSII und pLACZi-BSII	pGAD424 pGC1 pGC1 Δ DIM1	diese Arbeit
HIS-BSII _{m1}	pHISi-BSII _{m1}	pGAD424 pGC1	diese Arbeit
HIS-BSII _{m2}	pHISi-BSII _{m2}	pGAD424 pGC1	diese Arbeit
HIS-BSII _{m3}	pHISi-BSII _{m3}	pGAD424 pGC1	diese Arbeit
LACZ-BSII _{m1}	pLACZi-BSII _{m1}	pGAD424 pGC1	diese Arbeit
LACZ-BSII _{m2}	pLACZi-BSII _{m2}	pGAD424 pGC1	diese Arbeit
LACZ-BSII _{m3}	pLACZi-BSII _{m3}	pGAD424 pGC1	diese Arbeit

Tab. 6: DNA-Sequenz des isolierten cDNA-Klons (*cpcR1*) von *Acremonium chrysogenum*. Die angegebene Sequenz umfaßt 2546 bp (SEQ ID NO.: 20) und ist als einzelsträngiges, lineares DNA-Molekül in 5'-3'-Richtung dargestellt.

10	20	30	40	50	60
TCCAATATGC	CTGCTATGTA	CGGACAAGGG	TGGTCATCAT	CGTTTTCCCA	GCAGAACCAC
70	80	90	100	110	120
GCCGGATATT	CCACGCAGCC	GCCGCAGCCG	AGCCAGCAAG	GACAGACATC	GCAGATCCTC
130	140	150	160	170	180
CTTCTGCAGC	AAGCAAGCAA	CCCCGATGCG	GCATCAAATC	ATTTCCCTAT	GAACGGCATG
190	200	210	220	230	240
AACGGCCACC	CAATGGAATC	CCAATTTTCAT	GCTCACCCGA	TGCCAGCACA	ACACCCGATA
250	260	270	280	290	300
CAACACTCGC	AGCCTCACCT	ACACCAGATG	CATGCTCAGC	ATCAAATGCC	GGCCCATGCG
310	320	330	340	350	360
TTGAACCCAC	ACACCTGGGC	CAACGGAGAG	AGCCAATCTG	GAATGGATCA	AACCGGTGAT
370	380	390	400	410	420
GCCCTGAAGG	ACAATGCGAA	TGGTCTCAA	AAGGGAAAGT	CTAAGAGGAG	GTCGAATGCC
430	440	450	460	470	480
AACAATGACG	GCGAGATGAA	GGAGCTTTTC	GAGAGCAACC	GCCACAAGTC	GTTGGAGGAG
490	500	510	520	530	540
ATTGTTTCAGG	ATGTCAGGGG	CAACGAGCGC	GGTCCCAACG	CAGAGCGCAA	ACGTCAACTA
550	560	570	580	590	600
TATGCCATGC	TCTGGCTCAA	CTCAGCGTGC	GTGGAAAGCA	AGAACTCGAT	CCCACGCGGC
610	620	630	640	650	660
CGGGTGTATC	ACACTTATGC	CTCAAAATGT	ACCGACAGCA	GAGTCGTAGT	TCTCAACCCT
670	680	690	700	710	720
GCTAGTTTCG	GGAAGTTGGT	GCGAGTCGTC	TTCCCATCCA	TCAAAACCCG	ACGACTTGGA
730	740	750	760	770	780
GTCCGGGGCG	AGTCAAAGTA	TCACTATTGC	AACTTTCAAC	TCCGTGATCC	GCCACCGCCC
790	800	810	820	830	840
GAGGTGAGCG	GGCTGCGAGA	CGCTAGCGTT	CCCGCTCCCG	AAGAGGCAGC	TAAGGGGGAG
850	860	870	880	890	900
GAGTTTGACT	TCAACACACC	GCCGAACCAG	CAAATGACG	TCAAAGGAGC	CAGTCGTTTA
910	920	930	940	950	960
CCATCGCCGG	AGGATGCTTC	TCATCCACCG	GTCTCCCGCG	CCACATCACG	GACAAGCCAT
970	980	990	1000	1010	1020
GGATTGGGTT	CGCACAGCCG	TTACATTGTT	CCGCAGCCGC	GGCGACTGAA	ATCGGGATGG
1030	1040	1050	1060	1070	1080

GCTGGGGGCA TCTGCGCAAC GTCCAAGACG CGCATCAAGC TCCGGGTCGC AACAGAGGCG
 1090 1100 1110 1120 1130 1140
 GAAACGAAAT TCGATTGGGA CGAGCCACTG GTGTTGCCGC CGATCGACCC GTTCCTGCCG
 1150 1160 1170 1180 1190 1200
 CCGCACACAG ATAAGGACGC GGCCGCCTCG CTGGTGGCCC TGTACCGCGC GCAGTTGACA
 1210 1220 1230 1240 1250 1260
 TCCCTCGCTG AGGCCTTCCG ATACATCAGG GACAAGAGCT TTTTCCACCT GTACACGTCC
 1270 1280 1290 1300 1310 1320
 TTCCACGGCA CATTGACGAT GCCTGTACAG AAGCTCTTTG CGCATCCGAG CATTGCTCCT
 1330 1340 1350 1360 1370 1380
 TGGATCGAGG AGTGCGACCT TGTTCTGTAC CAGCGGCTGA CCAAATTCGC CTTCAACATG
 1390 1400 1410 1420 1430 1440
 TCGCTCGTGG TGGTGCCTAA TGTGTATCTC AACCGGATGC GGTCCATCGC CGACCGTCTC
 1450 1460 1470 1480 1490 1500
 GTGCCGCACA TTGTGGACGT GTTCGGCGGG CATCCTGAGC ATGTGGTGCA AGCCAAGGCG
 1510 9 1520 1530 1540 1550 1560
 GGTCCGGCGG CCATCTTCGT AGGCATCCTG GAGCGGATGA CGCGTGTGAA CAAGACGGCG
 1570 1580 1590 1600 1610 1620
 CACGCGGCGG CCAGACCACT GGCCTTGGAC GCCAACAGGG ACCAGATGTA CGCGGATTGG
 1630 1640 1650 1660 1670 1680
 CTGGAAGTGG TCAACGCGCG CAAGATTGCC GAGTGCGTGC CGACGCGGGG AATGGACGAT
 1690 1700 1710 1720 1730 1740
 GTGGCGGAGC TACTTGTCAG GGAGATGCGG TATCTCGTGG ACCCCAAGAA CTATATTCCG
 1750 1760 1770 1780 1790 1800
 GATGACGAGA CTTCTGAATGC CGATGCTGGA GGGGCCCGGT CACCGTTTGC CGTCAACCGA
 1810 1820 1830 1840 1850 1860
 TGGCGGGACT TTCTGATGAG TCTGCCCCGC AAGTTCCCGT ATGCGTCACA CGAAGACATT
 1870 1880 1890 1900 1910 1920
 GTTTGGTGCG TGGAGAGGGT GGGGACGGCG ATTGTGAGGG AGCTCACCTT GCAAGGGGGC
 1930 1940 1950 1960 1970 1980
 ACCAGCTTCA CAACGTGGTG GTCTATCAAG ACGTTCCTCG ACGAGGAGAT TATGTACCTG
 1990 2000 2010 2020 2030 2040
 GCCGAGGTTG GCGGGCTCAT GCATACGAAG AGCCTGCGCG ATGAGCCGTC AAGGCAGCAA
 2050 2060 2070 2080 2090 2100
 GCGACAACGG TCGTGACGGA AAGAGGCCAA AAAGGGGCCA CAGCGGACGA CGACGGTGAG
 2110 2120 2130 2140 2150 2160
 GTGGTGATGG GCGACGCAGG TCAGGAGGAA CCGATGGCGG TGGCGGCGTC GCCGGTGAC
 2170 2180 2190 2200 2210 2220
 ACAGACCGGG CGCCATTCCC ACCGAAACAG GGACATGCGG CCAACATGGG CGACGCGGCA
 2230 2240 2250 2260 2270 2280

```

AGTGATGCGC ACGATGACAG TGGGATAGGG ATCCGGACCC CCGAGACGGA CTTTCCGGTG
      2290      2300      2310      2320      2330      2340
GACAAGTTTG GACTCCATGA GGCGGAGGCG GGGGTCACAG AACTAGCCTA CGAAGGGTTG
      2350      2360      2370      2380      2390      2400
GATCGAGAAT GGGCGGCTCC TGCATGACGA GGCCTGACGA AGGCACGAAG AGATGAACAG
      2410      2420      2430      2440      2450      2460
CACGTGTTGG ATTAACGGTA TATCCCGGGG TTCTGGTTAT GGGTCCGGAG TAGGTCATGT
      2470      2480      2490      2500      2510      2520
CACTAAGGGG TCTATTGGGT AAATGTGATG TTGGGTCACT CTTCGAATTA TACCTCATTA
      2530      2540      2550      2560      2570      2580
ATGCCACTAC TGCCTGATGA TAAAAA

```

Tab. 7: Aminosäuresequenz, abgeleitet vom ORF 786 aus dem cDNA-Klon (*cpcR1*) von *Acremonium chrysogenum*. Der offenen Leserahmen kodiert für 786 Aminosäuren (SEQ ID NO.: 21).

Met	Arg	Ala	Asp	Glu	Pro	Ala	Ala	Ala	Ile	Gln	Ala	Ser	Ser	Ala	Ala	Met	Arg	Pro	Gln	10	20
Ser	Arg	Ala	Ser	Thr	Ser	Ser	Thr	His	Ser	Ala	Ala	Thr	Gln	Pro	Thr	Thr	Asp	Ala	His	30	40
Asp	Tyr	Ser	Asn	Met	Pro	Ala	Met	Tyr	Gly	Gln	Gly	Trp	Ser	Ser	Ser	Phe	Ser	Gln	Gln	50	60
Asn	His	Ala	Gly	Tyr	Ser	Thr	Gln	Pro	Pro	Gln	Pro	Ser	Gln	Gln	Gly	Gln	Thr	Ser	Gln	70	80
Ile	Leu	Leu	Leu	Gln	Gln	Ala	Ser	Asn	Pro	Asp	Ala	Ala	Ser	Asn	His	Phe	Pro	Met	Asn	90	100
Gly	Met	Asn	Gly	His	Pro	Met	Glu	Ser	Gln	Phe	His	Ala	His	Pro	Met	Pro	Ala	Gln	His	110	120
Pro	Ile	Gln	His	Ser	Gln	Pro	His	Leu	His	Gln	Met	His	Ala	Gln	His	Gln	Met	Pro	Ala	130	140
His	Ala	Leu	Asn	Pro	His	Thr	Trp	Ala	Asn	Gly	Glu	Ser	Gln	Ser	Gly	Met	Asp	Gln	Thr	150	160
Gly	Asp	Ala	Leu	Lys	Asp	Asn	Ala	Asn	Gly	Leu	Lys	Lys	Gly	Lys	Ser	Lys	Arg	Arg	Ser	170	180
Asn	Ala	Asn	Asn	Asp	Gly	Glu	Met	Lys	Glu	Leu	Phe	Glu	Ser	Asn	Arg	His	Lys	Ser	Leu	190	200
Glu	Glu	Ile	Val	Gln	Asp	Val	Arg	Gly	Asn	Glu	Arg	Gly	Pro	Asn	Ala	Glu	Arg	Lys	Arg	210	220
Gln	Leu	Tyr	Ala	Met	Leu	Trp	Leu	Asn	Ser	Ala	Cys	Val	Glu	Ser	Lys	Asn	Ser	Ile	Pro	230	240
Arg	Gly	Arg	Val	Tyr	His	Thr	Tyr	Ala	Ser	Lys	Cys	Thr	Asp	Asp	Arg	Val	Val	Val	Leu	250	260
Asn	Pro	Ala	Ser	Phe	Gly	Lys	Leu	Val	Arg	Val	Val	Phe	Pro	Ser	Ile	Lys	Thr	Arg	Arg	270	280
Leu	Gly	Val	Arg	Gly	Glu	Ser	Lys	Tyr	His	Tyr	Cys	Asn	Phe	Gln	Leu	Arg	Asp	Pro	Pro	290	300
Pro	Pro	Glu	Val	Ser	Gly	Leu	Arg	Asp	Ala	Ser	Val	Pro	Ala	Pro	Glu	Glu	Ala	Ala	Lys	310	320
Gly	Glu	Glu	Phe	Asp	Phe	Asn	Thr	Pro	Pro	Asn	Gln	Gln	Asn	Asp	Val	Lys	Gly	Ala	Ser	330	340
Arg	Leu	Pro	Ser	Pro	Glu	Asp	Ala	Ser	His	Pro	Pro	Val	Ser	Arg	Ala	Thr	Ser	Arg	Thr	350	360
																				370	380

Ser	His	Gly	Leu	Gly	Ser	His	Ser	Arg	Tyr	Ile	Val	Pro	Gln	Pro	Arg	Arg	Leu	Lys	Ser	
									390											400
Gly	Trp	Ala	Gly	Gly	Ile	Cys	Ala	Thr	Ser	Lys	Thr	Arg	Ile	Lys	Leu	Arg	Val	Ala	Thr	
									410											420
Glu	Ala	Glu	Thr	Lys	Phe	Asp	Trp	Asp	Glu	Pro	Leu	Val	Leu	Pro	Pro	Ile	Asp	Pro	Phe	
									430											440
Leu	Pro	Pro	His	Thr	Asp	Lys	Asp	Ala	Ala	Ala	Ser	Leu	Val	Ala	Leu	Tyr	Arg	Ala	Gln	
									450											460
Leu	Thr	Ser	Leu	Ala	Glu	Ala	Phe	Arg	Tyr	Ile	Arg	Asp	Lys	Ser	Phe	Phe	His	Leu	Tyr	
									470											480
Thr	Ser	Phe	His	Gly	Thr	Leu	Thr	Met	Pro	Val	Gln	Lys	Leu	Phe	Ala	His	Pro	Ser	Ile	
									490											500
Ala	Pro	Trp	Ile	Glu	Glu	Cys	Asp	Leu	Val	Leu	Tyr	Gln	Arg	Leu	Thr	Lys	Phe	Ala	Phe	
									510											520
Thr	Met	Ser	Leu	Val	Val	Val	Pro	Asn	Val	Tyr	Leu	Asn	Arg	Met	Arg	Ser	Ile	Ala	Asp	
									530											540
Arg	Leu	Val	Pro	His	Ile	Val	Asp	Val	Phe	Gly	Gly	His	Pro	Glu	His	Val	Val	Gln	Ala	
									550											560
Lys	Ala	Gly	Pro	Ala	Ala	Ile	Phe	Val	Gly	Ile	Leu	Glu	Arg	Met	Thr	Arg	Val	Asn	Lys	
									570											580
Thr	Ala	His	Ala	Ala	Ala	Arg	Pro	Val	Ala	Leu	Asp	Ala	Asn	Arg	Asp	Gln	Met	Tyr	Ala	
									590											600
Asp	Trp	Leu	Glu	Leu	Val	Asn	Ala	Arg	Lys	Ile	Ala	Glu	Cys	Val	Pro	Thr	Arg	Gly	Met	
									610											620
Asp	Asp	Val	Ala	Glu	Leu	Leu	Val	Arg	Glu	Met	Arg	Tyr	Leu	Val	Asp	Pro	Lys	Asn	Tyr	
									630											640
Ile	Pro	Asp	Asp	Glu	Thr	Ser	Asn	Ala	Asp	Ala	Gly	Gly	Ala	Arg	Ser	Pro	Phe	Ala	Val	
									650											660
Asn	Arg	Trp	Arg	Asp	Phe	Leu	Met	Ser	Leu	Pro	Gly	Lys	Phe	Pro	Tyr	Ala	Ser	His	Glu	
									670											680
Asp	Ile	Val	Trp	Cys	Val	Glu	Arg	Val	Gly	Thr	Ala	Ile	Val	Arg	Glu	Leu	Thr	Leu	Gln	
									690											700
Gly	Gly	Thr	Ser	Phe	Thr	Thr	Trp	Trp	Ser	Ile	Lys	Thr	Phe	Leu	Asp	Glu	Glu	Ile	Met	
									710											720
Tyr	Leu	Ala	Glu	Val	Gly	Gly	Leu	Met	His	Thr	Lys	Ser	Leu	Arg	Asp	Glu	Pro	Ser	Arg	
									730											740
Gln	Gln	Ala	Thr	Thr	Val	Val	Thr	Glu	Arg	Gly	Gln	Lys	Gly	Ala	Thr	Ala	Asp	Asp	Asp	
									750											760
Gly	Glu	Val	Val	Met	Gly	Asp	Ala	Gly	Gln	Glu	Glu	Pro	Met	Ala	Val	Ala	Ala	Ser	Pro	
									770											780
Val	His	Thr	Asp	Arg	Ala	Pro	Phe	Pro	Pro	Lys	Gln	Gly	His	Ala	Ala	Asn	Met	Gly	Asp	
									790											800
Ala	Ala	Ser	Asp	Ala	His	Asp	Asp	Ser	Gly	Ile	Gly	Ile	Arg	Thr	Pro	Glu	Thr	Asp	Phe	
									810											820
Pro	Val	Asp	Lys	Phe	Gly	Leu	His	Glu	Ala	Glu	Ala	Gly	Val	Thr	Glu	Leu	Ala	Tyr	Glu	
									830											840
Gly	Leu	Asp	Arg	Glu	Trp	Ala	Ala	Pro	Ala											

Tab. 8: Aminosäuresequenz-Vergleich der DNA-Bindedomänen einiger Mitglieder der RFX-Proteinfamilie. Identische Aminosäuren sind durch graue Boxen hervorgehoben. In der Konsensus-Sequenz wurden auch solche Positionen berücksichtigt, an denen 2 verschiedene Aminosäuren auftreten. Abkürzungen: AcCPCR1: *Acremonium chrysogenum* CPR1-Protein (SEQ ID NO.: 22), SpSAK1: *Schizosaccharomyces pombe* SAK1-Protein (U19978) (SEQ ID NO.: 23), ScRFX: *Saccharomyces cerevisiae* RFX-Protein (U17246) (SEQ ID NO.: 24), HsRFX1: *Homo sapiens* RFX1-Protein (A20492) (SEQ ID NO.: 25). In Klammern sind die „accession numbers“ der EMBL Datenbank (Heidelberg) angegeben.

	1	15	30	45	60	75																																																																		
1	ACPCR1	AM	W	NSACVESKNS-IP	GRVYHTY	ASKG	TD	SRVV	VLN	PAS	F	GR	V	RV	V	PS	IK	RR	TR	TC	VR	GE	SK	YH	C	N	F	Q	L	R	D																																									
2	SPSAK1	GE	W	L	K	R	A	C	E	E	Q	Q	D	A	A	V	Q	R	N	Q	I	Y	A	H	I	V	E	I	C	N	S	L	H	I	K	P	L	N	S	A	S	F	G	R	V	L	L	E	P	S	I	K	T	R	R	I	C	M	R	G	H	S	K	V	H	C	G	I	K	I	R	G
3	SCREFX	ALL	W	L	M	K	N	C	K	W	Q	H	D	S	Y	V	P	R	G	K	I	F	A	Q	I	A	S	S	S	Q	N	N	L	K	P	L	S	Q	S	L	G	K	I	R	T	V	E	P	D	L	T	R	I	R	I	C	M	R	G	H	S	K	V	H	C	G	L	K	L	T	V	
4	HsRFX1	TVQ	W	L	D	N	Y	E	T	A	E	G	V	S	L	P	R	S	T	L	Y	C	H	L	L	H	S	E	Q	K	L	E	P	V	N	A	S	F	G	R	V	L	L	E	P	S	I	K	T	R	R	I	C	M	R	G	H	S	K	V	H	C	G	L	K	L	T	V				
cons.		WL		C		Y		Y		Y		Y		C		PLN	ASF	G	K	L	V	R	V	F	P	L	T	R	R	L	G	R	G	S	K	Y	H	Y	C	G	L	K	L	T	V	YN	I																									

Tab. 9: Aminosäuresequenz-Vergleich der Dimerisationsdomänen einiger Mitglieder der RFX-Familie. Graue Boxen markieren identische Aminosäuren, weiße Positionen mit 2 verschiedenen Aminosäuren. Abkürzungen: AcCPCR1: *Acromonium chrysogenum* CPCR1-Protein (SEQ ID NO.: 26), SpSAK1: *Schizosaccharomyces pombe* SAK1-Protein (U19978) (SEQ ID NO.: 27), ScRFX: *Saccharomyces cerevisiae* RFX-Protein (U17246) (SEQ ID NO.: 28), HsRFX1: *Homo sapiens* RFX1-Protein (A20492) (SEQ ID NO.: 29). In Klammern sind die „accession numbers“ der EMBL Datenbank (Heidelberg) angegeben.

	1	15	30	45	60	75	90
1	ACCPCR1	IFVGILRMTTRVNTAKHAAARPVALDANRDQMAAEWLELNNARKIAE-----CVPTRG---MDDV---AELL---VREMRVIVDP-KNYIPD					
2	SpSAK1	IFSNILSRLLRVNDTAKHAAARFLANPADRHILICNDWERFSTRFIVHRELMCNDKEAVAALDEW---YSILSTCSNPSTLIDPLKD---K					
3	ScRFX	NETKLVKALLIKLLKELILNELK--SFHIFKSGMNNEMKNIINLDDILEMINE--DDTNSETNTIMQHLQGFQCVFVTKHINSSMSV-----					
4	HsRFX1	MAAQTLRHYTSLNHLAQAAARAVLQNTAQINQMLSDINR--DFANVDEQASWV-CRCEDRVVQRLE--QDFKVTLQQQNSI---EQWAAW-					
	cons.	F L R VN A AA A M DW V I V					L
		V K LL I NF I I L V					
	91	105	120	135	150	165	
1	ACCPCR1	DETSNA--DAGGARSPPAV-NRWRDFLMSILPGKRPYASHEDIV-WVERVGTAINVRELTIQGGTSFTTWWSIKTFIDEIMYLAE					
2	SpSAK1	HEASDTSMKRVELRQIDGVLDRMADFFLELPSPRPSCSPRMFL-LLGALQTSVLREITVSGGEAFGALWVIRCWVDEYMTWVAE					
3	ScRFX	--SNDPSVSIECKSLNE--MIKDFCSFISLQSKS--SCLKLID--STRFRNAILIGDISIKSNENLLSWLFLANNVMGQLLNVCFF					
4	HsRFX1	---LDGVVSQVL-----KP-----YQGS--AGEPKAAKLFLLKWSF--YSSMVI RDLTIRSAASEGSEFHILIRLLLYDEYMYVLI					
	cons.	D N	SL FP S C I RELTL F I DE Y E				W
			E-	- A S V GDISV L L GQ			

Tab. 10: Nukleotidsequenzen der Vektoren und Angabe rekombinanter Literaturreferenzen

Vektor	bakterieller Anteil	5' pcbC	Leaderpeptid	Linker	3' pcbC
pIR11	Stratagene, Heidelberg	Menne et al. 1994	-	Linker A	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233		Tab. 11	Pos. 1021-1462
	Pos. 728-2961				
pIR12	Stratagene, Heidelberg	Menne et al. 1994	-	Linker B	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233		Tab. 11	Pos. 1021-1462
	Pos. 728-2961				
pIR13	Stratagene, Heidelberg	Menne et al. 1994	-	Linker C	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233		Tab. 11	Pos. 1021-1462
	Pos. 728-2961				
pIR21	Stratagene, Heidelberg	Menne et al. 1994	von Fusarium sp.	Linker A	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233	42 bp, Tab. 12 B	Tab. 11	Pos. 1021-1462
	Pos. 728-2961				
pIR22	Stratagene, Heidelberg	Menne et al. 1994	von Fusarium sp.	Linker B	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233	42 bp, Tab. 12 B	Tab. 11	Pos. 1021-1462

Vektor	bakterieller Anteil	5' pcbC	Leaderpeptid	Linker	3' pcbC
pIR23	Pos. 728-2961				
	Strategene, Heidelberg	Menne et al. 1994	von Fusarium sp.	Linker C	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233	42 bp, Tab. 12 B	Tab. 11	Pos. 1021-1462
pIR31	Pos. 728-2961				
	Strategene, Heidelberg	Menne et al. 1994	von Acremonium	Linker A	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233	chrysogenum	Tab. 11	Pos. 1021-1462
pIR32	Pos. 728-2961				
	Strategene, Heidelberg	Menne et al. 1994	60 bp, Tab. 12 C	Linker B	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233	chrysogenum	Tab. 11	Pos. 1021-1462
pIR33	Pos. 728-2961				
	Strategene, Heidelberg	Menne et al. 1994	60 bp, Tab. 12 C	Linker C	Samson et al. 1985
	Pos. 1-659	Pos. 7-1233	chrysogenum	Tab. 11	Pos. 1021-1462
	Pos. 728-2961				
			60 bp, Tab. 12 C		

Tab. 11: Die charakteristischen Klonierungsstellen der neun Vektoren und die davon abgeleiteten Restriktionsschnittstellen, welche für Klonierungen geeignet sind. Zu beachten ist, daß sich die drei Klonierungsstellen nur durch ein bzw. zwei Basenpaare unterscheiden.

1. Leseraster (Oligonukleotide Nr. 1+2)

	HindIII	NruI	EcoRI	
A	5' C ATG AAG CTT TCG CGA GAA TCC 3' (SEQ ID NO.: 30)			PIR22,21,31
	3' TTC GAA AGC GCT CTT AAG 5' (SEQ ID NO.: 31)			

2. Leseraster (Oligonukleotide Nr. 3 + 4)

B	5' C ATG <u>CAA</u> GCT TTC GCG AGA ATT C 3' (SEQ ID NO.: 32)	PIR12,22,32
	3' GTT CGA AAG CGC TCT TAA G 5' (SEQ ID NO.: 33)	

3. Leseraster (Oligonukleotide Nr. 5 + 6)

C	5' C ATG <u>CTA</u> AGC TTT CGC GAG AAT TC 3' (SEQ ID NO.: 34)	PIR13,23,33
	3' <u>GAT</u> TCG AAA GCG CTC TTA AG 5' (SEQ ID NO.: 35)	

Tab. 12: Übergangssequenzen im Bereich der multiplen Klonierungsstelle (eingerahmt) in den angegebenen Expressionsvektoren. In den Sequenzen wurden solche Nukleotide unterstrichen, die gegenüber dem erstgenannten Plasmid verschieden sind.

A. Vektoren ohne Signalsequenz

pIR11

1. Leseraster

		HindIII	NruI	EcoRI	
AAACCGTCACC	ATG	AAG CTT	TCG CGA	GAA TTC	GGGTCGATCAGG (SEQ ID NO.: 36)
TTTGGCAGTGG	TAC	TTC GAA	AGC GCT	CTT AAG	CCCAGCTAGTCC (SEQ ID NO.: 37)

pIR12

2. Leseraster

AAACCGTCACC	ATG	CAA GCT	TTC GCG	AGA ATT	CGGGTCGATCAGG (SEQ ID NO.: 38)
TTTGGCAGTGG	TAC	GTT CGA	AAG CGC	TCT TAA	GCCCAGCTAGTCC (SEQ ID NO.: 39)

pIR13

3. Leseraster

AAACCGTCACC	ATG	CTA AGC	TTT CGC	GAG AAT	TOGGGTCGATCAGG (SEQ ID NO.: 40)
TTTGGCAGTGG	TAC	GAT TCG	AAA GCG	CTC TTA	AGCCCAGCTAGTCC (SEQ ID NO.: 41)

B. Vektoren mit Signalsequenz aus *Fusarium*

pIR21

1. Leseraster

AAACCGTCACC ATG CGT CTG TCC ATC ATC GCA GTC CTT CCC CTG GCC CTC GCC
 TTTGGCAGTGG TAC GCA GAC AGG TAG TAG CGT CAG GAA GGG GAC CGG GAG CGG

	HindIII	NruI	EcoRI	
ATG	AAG CTT	TCG CGA	GAA TTC	GGGTCGATCAGG (SEQ ID NO.: 42)
TAC	TTC GAA	AGC GCT	CTT AAG	CCCAGCTAGTCC (SEQ ID NO.: 43)

pIR22

2. Leseraster

AAACCGTCACC ATG CGT CTG TCC ATC ATC GCA GTC CTT CCC CTG GCC CTC GCC
 TTTGGCAGTGG TAC GCA GAC AGG TAG TAG CGT CAG GAA GGG GAC CGG GAG CGG

ATG	CAA GCT	TTC GCG	AGA ATT	CGGGTCGATCAGG (SEQ ID NO.: 44)
TAC	GTT CGA	AAG CGC	TCT TAA	GCCCAGCTAGTCC (SEQ ID NO.: 45)

pIR23

3. Leseraster

AAACCGTCACC ATG CGT CTG TCC ATC ATC GCA GTC CTT CCC CTG GCC CTC GCC
 TTTGGCAGTGG TAC GCA GAC AGG TAG TAG CGT CAG GAA GGG GAC CGG GAG CGG

ATG	CTA AGC	TTT CGC	GAG AAT	TCGGGTCGATCAGG (SEQ ID NO.: 46)
TAC	GAT TGC	AAA GCG	CTC TTA	AGCCCAGCTAGTCC (SEQ ID NO.: 47)

C. Vektoren mit Signalsequenz aus *Acremonium chrysogenum*

pIR31

1. Leseraster

AAACCGTCACC ATG GTC ACC CTC CGC CGC CTC GCC GTC CTC CTC GGC
 TTTGGCAGTGG TAC CAG TGG GAG GCG GCG GAG CGG CAG GAG GAG CCG

	HindIII	NruI	EcoRI
GCC ATC CCC GCC GCC CTC GCC GCC ATG	AAG CTT	TCG CGA	GAA TTC
CGG TAG GGG CGG CGG GAG CGG CGG TAC	TTC GAA	AGC GCT	CTT AAG

GGGTCGATCAGG (SEQ ID NO.: 48)

CCCAGCTAGTCC (SEQ ID NO.: 49)

pIR32

2. Leseraster

AAACCGTCACC ATG GTC ACC CTC CGC CGC CTC GCC GTC CTC CTC GGC GCC ATC CCC GOC
 TTTGGCAGTGG TAC CAG TGG GAG GCG GCG GAG CGG CAG GAG GAG CCG CGG TAG GGG CGG

GCC CTC GCC GCC ATG	GAA GCT	TTC GCG	AGA ATT	GGGTCGATCAGG (SEQ ID NO.: 50)
CGG GAG CGG CGG TAC	GTT CGA	AAG CGC	TCT TAA	CCCAGCTAGTCC (SEQ ID NO.: 51)

pIR33

3. Leseraster

AAACCGTCACC ATG GTC ACC CTC CGC CGC CTC GCC GTC CTC CTC GGC GCC ATC CCC GCC
 TTTGGCAGTGG TAC CAG TGG GAG GCG GCG GAG CGG CAG GAG GAG CCG CGG TAG GGG CGG

GCC CTC GCC GCC ATG	CTA AGC	TTT CGC	GAG AAT	TCGGGTCGATCAGG (SEQ ID NO.: 52)
CGG GAG CGG CGG TAC	GAT TGC	AAA GCG	CTC TAA	AGCCCAGCTAGTCC (SEQ ID NO.: 53)

Tabelle 13: DNA-Sequenz der genomischen Kopie des cpcR1-Gens von *Acremonium chrysogenum* (SEQ ID NO.: 57)

10	20	30	40	50	60
AGAATGAGAG	TTATTACCTG	GGAATCGTTC	GTGCAGTGGT	CCCAGCCACA	TGCACAAGGT
70	80	90	100	110	120
TGAGAGGTTC	GCAGCAGGGA	TGCGGGGCAA	CGCCAGGAAC	TTTGACAAAT	GAGGGTGCGT
130	140	150	160	170	180
TCCTGCTTAC	CTAAGTTGGC	TCGTGCCCAC	AAACCGGGAG	TCGTTTCTCC	TCCCGTTGGC
190	200	210	220	230	240
CTCGCAACAG	AAAAACACAC	CGACCACAAA	CAACAGCCCG	CCTTCCGTCC	ATCGAGCAAC
250	260	270	280	290	300
CACTTGCCTA	GCCTTGCCCG	ACGCGGGTCA	TCCAGACAGA	CCTTGCGCGG	AACCGGACGC
310	320	330	340	350	360
ATCACGTCAC	ATTTTCTTTT	ATTTCTTTTC	CTCGACAACG	ACACGCCTTC	AGCGCGCCCT
370	380	390	400	410	420
GCGCACCCAT	CGCTGTCGCG	CTGCTGCGGA	GCAGTGTTTG	TTGTGAACTG	CGACCGACCC
430	440	450	460	470	480
CCGACAGGCA	GCTTTCCTGT	GAAGTGGACC	TCCACCAAAA	CATCCCCATC	ATGGACCAAG
490	500	510	520	530	540
GTTTGTCCCA	CCCTGCTCCG	ATGGCAGATG	ACTTACTAGA	ACGGCTTGTT	GACAATGCGA
550	560	570	580	590	600
GCAGACGAGC	CGGCCGCCGC	GATTCAGGCG	TCTTCCGCCG	CGATGCGGCC	GCAATCTAGG
610	620	630	640	650	660
GCCTCGACCT	CCTCGACACA	CAGCGCCGCG	ACGCAACCCA	CCACAGATGC	CCATGATTAC
670	680	690	700	710	720
TCCAATATGC	CTGCTATGTA	CGGACAAGGG	TGGTCATCAT	CGTTTTCCCA	GCAGAACCAC

730 740 750 760 770 780
GCCGGATATT CCACGCAGCC GCCGCAGCCG AGCCAGCAAG GACAGACATC GCAGATCCTC

790 800 810 820 830 840
CTTCTGCAGC AAGCAAGCAA CCCCAGATGCG GCATCAAATC ATTTCCCTAT GAACGGCATG

850 860 870 880 890 900
AACGGCCACC CAATGGAATC CCAATTTTCAT GCTCACCCGA TGCCAGCACA ACACCCGATA

910 920 930 940 950 960
CAACACTCGC AGCCTCACCT ACACCAGATG CATGCTCAGC ATCAAATGCC GGCCCATGCG

970 980 990 1000 1010 1020
TTGAACCCAC ACACCTGGGC CAACGGAGAG AGCCAATCTG GAATGGATCA AACCGGTGAT

1030 1040 1050 1060 1070 1080
GCCCTGAAGG ACAATGCGAA TGGTCTCAAA AAGGGAAAGT CTAAGAGGAG GTCGAATGCC

1090 1100 1110 1120 1130 1140
ACAATGACG GCGAGATGAA GGAGCTTTTC GAGAGCAACC GCCACAAGTC GTTGGAGGAG

1150 1160 1170 1180 1190 1200
ATTGTTTCAGG ATGTCAGGGG CAACGAGCGC GGTCCCAACG CAGAGCGCAA ACGTCAACTA

1210 1220 1230 1240 1250 1260
TATGCCATGC TCTGGTAAGA ACGACAGAGA GACTGACCAG CGCTTCACGT TCACTGACAT

1270 1280 1290 1300 1310 1320
GCGACAGGCT CAACTCAGCG TGCGTGGAAA GCAAGAACTC GATCCCACGC GGCCGGGTGT

1330 1340 1350 1360 1370 1380
ATCACACTTA TGCCTCAAAA TGTACCGACG ACAGAGTCGT AGTTCTCAAC CCTGCTAGTT

1390 1400 1410 1420 1430 1440
TCGGGAAGTT GGTGCGAGTC GTCTTCCCAT CCATCAAAAC CCGACGACTT GGAGTCCGGG

1450 1460 1470 1480 1490 1500
GCGAGTCAAA GTATCACTAT TGCAACTTTC AACTCCGTGA TCCGCCACCG CCCGAGGTGA

1510 1520 1530 1540 1550 1560

GCGGGCTGCG AGACGCTAGC GTTCCCGCTC CCGAAGAGGC AGCTAAGGGG GAGGAGTTTG

1570 1580 1590 1600 1610 1620
ACTTCAAGTA AGTACCTATT GACAGGATGA TGCTGTATGC GCCTCTAACG TGTACCCAGC

1630 1640 1650 1660 1670 1680
ACACCGCCGA ACCAGCAAAA TGACGTCAAA GGAGCCAGTC GTTTACCATC GCCGGAGGAT

1690 1700 1710 1720 1730 1740
GCTTCTCATC CACCGGTCTC CCGCGCCACA TCACGGACAA GCCATGGATT GGGTTCGCAC

1750 1760 1770 1780 1790 1800
AGCCGTTACA TTGTTCCGCA GCCGCGGCGA CTGAAATCGG GATGGGCTGG GGGCATCTGC

1810 1820 1830 1840 1850 1860
GCAACGTCCA AGACGCGCAT CAAGCTCCGG GTCGCAACAG AGGCGGAAAC GAAATTCGAT

1870 1880 1890 1900 1910 1920
TGGGACGAGC CACTGGTGTT GCCGCCGATC GACCCGTTCC TGCCGCCGCA CACAGATAAG

1930 1940 1950 1960 1970 1980
GACGCGGCCG CCTCGCTGGT GGCCCTGTAC CGCGCGCAGT TGACATCCCT CGCTGAGGCC

1990 2000 2010 2020 2030 2040
TTCCGATACA TCAGGGACAA GAGCTTTTTC CACCTGTACA CGTCCTTCCA CGGCACATTG

2050 2060 2070 2080 2090 2100
ACGATGCCTG TACAGAAGCT CTTTGCGCAT CCGAGCATTG CTCCTTGAT CGAGGAGTGC

2110 2120 2130 2140 2150 2160
GACCTTGTTT TGTACCAGCG GCTGACCAAA TTCGCCTTCA CCATGTCGCT CGTGGTGGTG

2170 2180 2190 2200 2210 2220
CCTAATGTGT ATCTCAACCG GATGCGGTCC ATCGCCGACC GTCTCGTGCC GCACATTGTG

2230 2240 2250 2260 2270 2280
GACGTGTTCG GCGGGCATCC TGAGCATGTG GTGCAAGCCA AGGCGGGTCC GGCGGCCATC

2290 2300 2310 2320 2330 2340
TTCGTAGGCA TCCTGGAGCG GATGACGCGT GTGAACAAGA CGGCGCACGC GGCGGCCAGA

2350 2360 2370 2380 2390 2400
CCAGTGGCGT TGGACGCCAA CAGGGACCAG ATGTACGCGG ATTGGCTGGA ACTGGTCAAC

2410 2420 2430 2440 2450 2460
GCGCGCAAGA TTGCCGAGTG CGTGCCGACG CGGGGAATGG ACGATGTGGC GGAGCTACTT

2470 2480 2490 2500 2510 2520
GTCAGGGAGA TGCGGTATCT CGTGGACCCC AAGAACTATA TTCCGGATGA CGAGACTTCG

2530 2540 2550 2560 2570 2580
AATGCCGATG CTGGAGGGGC CCGGTCACCG TTTGCCGTCA ACCGATGGCG GGA CTTTCTG

2590 2600 2610 2620 2630 2640
ATGAGTCTGC CCGGCAAGTT CCCGTATGCG TCACACGAAG ACATTGTTTG GTGCGTGGAG

2650 2660 2670 2680 2690 2700
AGGGTGGGGA CGGCGATTGT GAGGGAGCTC ACCTTGCAAG GGGGCACCAG CTT CACAACG

2710 2720 2730 2740 2750 2760
TGGTGGTCTA TCAAGACGTT CCTCGACGAG GAGATTATGT ACCTGGCCGA GGT TGGCGGG

2770 2780 2790 2800 2810 2820
CTCATGCATA CGAAGAGCCT GCGCGATGAG CCGTCAAGGC AGCAAGCGAC AACGGTCGTG

2830 2840 2850 2860 2870 2880
ACGGAAAGAG GCCAAAAGG GGCCACAGCG GACGACGACG GTGAGGTGGT GATGGGCGAC

2890 2900 2910 2920 2930 2940
GCAGGTCAGG AGGAACCGAT GGCGGTGGCG GCGTCGCCGG TGCACACAGA CCGGGCGCCA

2950 2960 2970 2980 2990 3000
TTCCCACCGA AACAGGGACA TGCGGCCAAC ATGGGCGACG CGGCAAGTGA TGCGCACGAT

3010 3020 3030 3040 3050 3060
GACAGTGGGA TAGGGATCCG GACCCCCGAG ACGGACTTTC CGGTGGACAA GTTTGGACTC

3070 3080 3090 3100 3110 3120
CATGAGGCGG AGGCGGGGGT CACAGAACTA GCCTACGAAG GGT TGGATCG AGAATGGGCG

3130 3140 3150 3160 3170 3180
GCTCCTGCAT GACGAGGCCT GACGAAGGCA CGAAGAGATG AACAGCACGT GTTGGATTAA

3190 3200 3210 3220 3230 3240
CGGTATATCC CGGGGTTCTG GTTATGGGTC CGGAGTAGGT CATGTCAC TA AGGGGTCTAT

3250 3260 3270 3280 3290 3300
TGGGTAAATG TGATGTTGGG TCACTCTTCG AATTATACCT CATTAATGCC ACTACTGCCT

3310 3320 3330 3340 3350 3360
GATGATACTC TTATGGGGGG GGGGGGGGGG TTGAGCTACT CAATGCTTAA GTGAATCGGA

3370 3380 3390 3400 3410 3420
GAGCACTTTG CCTTTCCAGC GGAGACTTTT TGTTGTCGGC CAGTGATGCT TGGTTGTAGA

3430 3440 3450 3460 3470 3480
TGGAAGTGGT GGTCATGACT CGGAGGAGGG TGTAGGGTAG CTGTACTTGC GGGTGCGTTG

3490 3500 3510 3520 3530 3540
TTACGAAGAG TATTTGGGCG GCGGAAGACC GGGCGGATGC CGAACCCGGA CGTGCGGTAA

3550 3560 3570 3580 3590 3600
CGGGGATGAG CAGCGTCAGT CGAGTCATCC ATGATTTCAG GTCCAGGGTC CAAGTCGCGT

3610 3620 3630 3640 3650 3660
ATTCGTATTT GTATTCCGGA TTACTGTTTG CCAGTATGGA GTTTACGGTA CTAAGGGAGA

3670 3680 3690 3700 3710 3720
TTATTTAGTC GAC

SEQUENZPROTOKOLL

(1) ALLGEMEINE INFORMATION:

(i) ANMELDER:

- (A) NAME: Hoechst Marion Roussel Deutschland GmbH
- (B) STRASSE: -
- (C) ORT: Frankfurt
- (D) BUNDESLAND: -
- (E) LAND: Deutschland
- (F) POSTLEITZAHL: 65926
- (G) TELEPHON: 069-305-3005
- (H) TELEFAX: 069-35-7175
- (I) TELEX: -

- (ii) ANMELDETITEL: Regulation des Isopenicillin
N-Synthetase-Gens (pcbC) durch das cpcR1-Genprodukt

- (iii) ANZAHL DER SEQUENZEN: 56

(iv) COMPUTER-LESBARE FORM:

- (A) DATENTRÄGER: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPA)

(2) INFORMATION ZU SEQ ID NO: 1:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 13 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

- (ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..13

- (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

TTGCCGGGCC AAT

13

(2) INFORMATION ZU SEQ ID NO: 2:

- (i) SEQUENZ CHARAKTERISTIKA:
(A) LÄNGE: 10 Basenpaare
(B) ART: Nukleinsäure
(C) STRANGFORM: Einzel
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
(A) NAME/SCHLÜSSEL: exon
(B) LAGE: 1..10

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

TTAGGCCTAA

10

(2) INFORMATION ZU SEQ ID NO: 3:

- (i) SEQUENZ CHARAKTERISTIKA:
(A) LÄNGE: 24 Basenpaare
(B) ART: Nukleinsäure
(C) STRANGFORM: Einzel
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
(A) NAME/SCHLÜSSEL: exon
(B) LAGE: 1..24

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 3:

GTTGCCGGGC CAATCCCTGA GCTT

24

(2) INFORMATION ZU SEQ ID NO: 4:

- (i) SEQUENZ CHARAKTERISTIKA:
(A) LÄNGE: 24 Basenpaare
(B) ART: Nukleinsäure
(C) STRANGFORM: Einzel
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
(A) NAME/SCHLÜSSEL: exon
(B) LAGE: 1..24

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 4:

GTTGCCGGGC CAATCCCTGA GCTT

24

(2) INFORMATION ZU SEQ ID NO: 5:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 24 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..24

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 5:

AAGCTCAGGG ATTGGCCCGG CAAC

24

(2) INFORMATION ZU SEQ ID NO: 6:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 24 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..24

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 6:

GTTGCCGGGG GTTACCCTGA GCTT

24

(2) INFORMATION ZU SEQ ID NO: 7:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 24 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..24

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 7:

AAGCTCAGGG TAACCCCCGG CAAC

24

(2) INFORMATION ZU SEQ ID NO: 8:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 24 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Einzel
 - (D) TOPOLOGIE: linear

- (ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..24

- (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 8:

GTTGCCGGGC GTATCCCTGA GCTT

24

(2) INFORMATION ZU SEQ ID NO: 9:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 24 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Einzel
 - (D) TOPOLOGIE: linear

- (ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..24

- (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 9:

AAGCTCAGGG ATACGCCCGG CAAC

24

(2) INFORMATION ZU SEQ ID NO: 10:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 24 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Einzel
 - (D) TOPOLOGIE: linear

- (ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..24

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 10:

CAACCCGGGC CAATCCCTGA GCTT

24

(2) INFORMATION ZU SEQ ID NO: 11:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 24 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..24

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 11:

AAGCTCAGGG ATTGGCCCGG GTTG

24

(2) INFORMATION ZU SEQ ID NO: 12:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 80 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..80

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 12:

AATTCGTTGC CGGGCCAATC CCTGAGCTTG TTGCCGGGCC AATCCCTGAG CTTGTTGCCG

60

GGCCAATCCC TGAGCTTCCC

80

(2) ANGABEN ZU SEQ ID NO: 13:

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 76 Basenpaare
- (B) ART: Nucleotid
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Genom-DNA

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..76

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 13:

GCAACGGCCC GGTAGGGAC TCGAACAACG GCCCGGTTAG GGACTCGAAC AACGGCCCGG 60
TTAGGGACTC GAAGGG 76

(2) INFORMATION ZU SEQ ID NO: 14:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 80 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Einzel
 - (D) TOPOLOGIE: linear

- (ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..80

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 14:

AATTCGTTGC CGGGGGTTAC CCTGAGCTTG TTGCCGGGGG TTACCCTGAG CTTGTTGCCG 60
GGGGTTACCC TGAGCTTCCC 80

(2) INFORMATION ZU SEQ ID NO: 15:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 76 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Einzel
 - (D) TOPOLOGIE: linear

- (ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..76

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 15:

GGGAAGCTCA GGGTAACCCC CGGCAACAAG CTCAGGGTAA CCCCCGGCAA CAAGCTCAGG 60
GTAACCCCCG GCAACG 76

(2) INFORMATION ZU SEQ ID NO: 16:

- (i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 80 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..80

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 16:

AATTCGTTGC CGGGCGTATC CCTGAGCTTG TTGCCGGGCG TATCCCTGAG CTTGTTGCCG 60
GGCGTATCCC TGAGCTTCCC 80

(2) INFORMATION ZU SEQ ID NO: 17:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 76 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..76

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 17:

GGGAAGCTCA GGGATACGCC CGGCAACAAG CTCAGGGATA CGCCCGGCAA CAAGCTCAGG 60
GATACGCCCCG GCAACG 76

(2) INFORMATION ZU SEQ ID NO: 18:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 80 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..80

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 18:

AATTCCAACC CGGGCCAATC CCTGAGCTTC AACCCGGGCC AATCCCTGAG CTTCAACCCG 60
GGCCAATCCC TGAGCTTCCC 80

(2) INFORMATION ZU SEQ ID NO: 19:

- (i) SEQUENZ CHARAKTERISTIKA:
(A) LÄNGE: 76 Basenpaare
(B) ART: Nukleinsäure
(C) STRANGFORM: Einzel
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
(A) NAME/SCHLÜSSEL: exon
(B) LAGE: 1..76

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 19:

GGGAAGCTCA GGGATTGGCC CGGGTTGAAG CTCAGGGATT GGCCCGGGTT GAAGCTCAGG 60
GATTGGCCCCG GGTGG 76

(2) ANGABEN ZU SEQ ID NO: 20:

- (i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 2546 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Genom-DNA

- (ix) MERKMAL:
(A) NAME/SCHLÜSSEL: exon
(B) LAGE: 1..2546

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 20:

TCCAATATGC CTGCTATGTA CGGACAAGGG TGGTCATCAT CGTTTTCCCA GCAGAACCAC 60
GCCGGATATT CCACGCAGCC GCCGCAGCCG AGCCAGCAAG GACAGACATC GCAGATCCTC 120
CTTCTGCAGC AAGCAAGCAA CCCCgatGCG GCATCAAATC ATTTCCCTAT GAACGGCATG 180
AACGGCCACC CAATGGAATC CCAATTTTCAT GCTCACCCGA TGCCAGCACA ACACCCGATA 240
CAACACTCGC AGCCTCACCT ACACCAGATG CATGCTCAGC ATCAAATGCC GGCCCATGCG 300
TTGAACCCAC ACACCTGGGC CAACGGAGAG AGCCAATCTG GAATGGATCA AACCGGTGAT 360
GCCCTGAAGG ACAATGCGAA TGGTCTCAAA AAGGGAAAGT CTAAGAGGAG GTCGAATGCC 420

AACAATGACG GCGAGATGAA GGAGCTTTTC GAGAGCAACC GCCACAAGTC GTTGGAGGAG	480
ATTGTTTCAGG ATGTCAGGGG CAACGAGCGC GGTCCCAACG CAGAGCGCAA ACGTCAACTA	540
TATGCCATGC TCTGGCTCAA CTCAGCGTGC GTGGAAAGCA AGAACTCGAT CCCACGCGGC	600
CGGGTGTATC AACTTTATGC CTCAAAATGT ACCGACAGCA GAGTCGTAGT TCTCAACCCT	660
GCTAGTTTCG GGAAGTTGGT GCGAGTCGTC TTCCCATCCA TCAAAACCCG ACGACTTGGA	720
GTCCGGGGCG AGTCAAAGTA TCACTATTGC AACTTTCAAC TCCGTGATCC GCCACCGCCC	780
GAGGTGAGCG GGCTGCGAGA CGCTAGCGTT CCCGCTCCCG AAGAGGCAGC TAAGGGGGAG	840
GAGTTTGA CTCAACACACC GCCGAACCAG CAAAATGACG TCAAAGGAGC CAGTCGTTTA	900
CCATCGCCGG AGGATGCTTC TCATCCACCG GTCTCCCGCG CCACATCAGC GACAAGCCAT	960
GGATTGGGTT CGCACAGCCG TTACATTGTT CCGCAGCCGC GGCGACTGAA ATCGGGATGG	1020
GCTGGGGGCA TCTGCGCAAC GTCCAAGACG CGCATCAAGC TCCGGGTCGC AACAGAGGCG	1080
GAAACGAAAT TCGATTGGGA CGAGCCACTG GTGTTGCCGC CGATCGACCC GTTCCTGCCG	1140
CCGCACACAG ATAAGGACGC GGCCGCCTCG CTGGTGGCCC TGTACCGCGC GCAGTTGACA	1200
TCCCTCGCTG AGGCCTTCCG ATACATCAGG GACAAGAGCT TTTTCCACCT GTACACGTCC	1260
TTCCACGGCA CATTGACGAT GCCTGTACAG AAGCTCTTTG CGCATCCGAG CATTGCTCCT	1320
TGGATCGAGG AGTGCGACCT TGTCTGTAC CAGCGGCTGA CCAAATTCGC CTTCAACCATG	1380
TCGCTCGTGG TGGTGCCTAA TGTGTATCTC AACCGGATGC GGTCCATCGC CGACCGTCTC	1440
GTGCCGCACA TTGTGGACGT GTTCGGCGGG CATCCTGAGC ATGTGGTGCA AGCCAAGGCG	1500
GGTCCGGCGG CCATCTTCGT AGGCATCCTG GAGCGGATGA CGCGTGTGAA CAAGACGGCG	1560
CACGCGGCGG CCAGACCAGT GGC GTTGGAC GCCAACAGGG ACCAGATGTA CGCGGATTGG	1620
CTGGAACTGG TCAACGCGCG CAAGATTGCC GAGTGCGTGC CGACGCGGGG AATGGACGAT	1680
GTGGCGGAGC TACTTGTCAG GGAGATGCGG TATCTCGTGG ACCCCAAGAA CTATATTCCG	1740
GATGACGAGA CTTCGAATGC CGATGCTGGA GGGGCCCGGT CACCGTTTGC CGTCAACCGA	1800
TGGCGGGACT TTCTGATGAG TCTGCCCCGGC AAGTTCCCGT ATGCGTCACA CGAAGACATT	1860
GTTTGGTGCG TGGAGAGGGT GGGGACGGCG ATTGTGAGGG AGCTCACCTT GCAAGGGGGC	1920
ACCAGCTTCA CAACGTGGTG GTCTATCAAG ACGTTCCTCG ACGAGGAGAT TATGTACCTG	1980
GCCGAGGTTG GCGGGCTCAT GCATACGAAG AGCCTGCGCG ATGAGCCGTC AAGGCAGCAA	2040
GCGACAACGG TCGTGACGGA AAGAGGCCAA AAAGGGGCCA CAGCGGACGA CGACGGTGAG	2100
GTGGTGATGG GCGACGCAGG TCAGGAGGAA CCGATGGCGG TGGCGGCGTC GCCGGTGAC	2160
ACAGACCGGG CGCCATTCCC ACCGAAACAG GGACATGCGG CCAACATGGG CGACGCGGCA	2220

```

AGTGATGCGC ACGATGACAG TGGGATAGGG ATCCGGACCC CCGAGACGGA CTTTCCGGTG      2280
GACAAGTTTG GACTCCATGA GCGGGAGGCG GGGGTCACAG AACTAGCCTA CGAAGGGTTG      2340
GATCGAGAAT GGGCGGCTCC TGCATGACGA GGCCTGACGA AGGCACGAAG AGATGAACAG      2400
CACGTGTTGG ATTAACGGTA TATCCCGGGG TTCTGGTTAT GGGTCCGGAG TAGGTCATGT      2460
CACTAAGGGG TCTATTGGGT AAATGTGATG TTGGGTCACT CTTCGAATTA TACCTCATTA      2520
ATGCCACTAC TGCCTGATGA TAAAAA                                           2546

```

(2) ANGABEN ZU SEQ ID NO: 21:

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 830 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: Peptide
- (B) LÄNGE: 1..830

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 21:

```

Met Arg Ala Asp Glu Pro Ala Ala Ala Ile Gln Ala Ser Ser Ala Ala
1           5           10           15
Met Arg Pro Gln Ser Arg Ala Ser Thr Ser Ser Thr His Ser Ala Ala
20           25           30
Thr Gln Pro Thr Thr Asp Ala His Asp Tyr Ser Asn Met Pro Ala Met
35           40           45
Tyr Gly Gln Gly Trp Ser Ser Ser Phe Ser Gln Gln Asn His Ala Gly
50           55           60
Tyr Ser Thr Gln Pro Pro Gln Pro Ser Gln Gln Gly Gln Thr Ser Gln
65           70           75           80
Ile Leu Leu Leu Gln Gln Ala Ser Asn Pro Asp Ala Ala Ser Asn His
85           90           95
Phe Pro Met Asn Gly Met Asn Gly His Pro Met Glu Ser Gln Phe His
100          105          110
Ala His Pro Met Pro Ala Gln His Pro Ile Gln His Ser Gln Pro His
115          120          125
Leu His Gln Met His Ala Gln His Gln Met Pro Ala His Ala Leu Asn
130          135          140
Pro His Thr Trp Ala Asn Gly Glu Ser Gln Ser Gly Met Asp Gln Thr
145          150          155          160

```


Gly Asp Ala Leu Lys Asp Asn Ala Asn Gly Leu Lys Lys Gly Lys Ser
 165 170 175
 Lys Arg Arg Ser Asn Ala Asn Asn Asp Gly Glu Met Lys Glu Leu Phe
 180 185 190
 Glu Ser Asn Arg His Lys Ser Leu Glu Glu Ile Val Gln Asp Val Arg
 195 200 205
 Gly Asn Glu Arg Gly Pro Asn Ala Glu Arg Lys Arg Gln Leu Tyr Ala
 210 215 220
 Met Leu Trp Leu Asn Ser Ala Cys Val Glu Ser Lys Asn Ser Ile Pro
 225 230 235 240
 Arg Gly Arg Val Tyr His Thr Tyr Ala Ser Lys Cys Thr Asp Asp Arg
 245 250 255
 Val Val Val Leu Asn Pro Ala Ser Phe Gly Lys Leu Val Arg Val Val
 260 265 270
 Phe Pro Ser Ile Lys Thr Arg Arg Leu Gly Val Arg Gly Glu Ser Lys
 275 280 285
 Tyr His Tyr Cys Asn Phe Gln Leu Arg Asp Pro Pro Pro Pro Glu Val
 290 295 300
 Ser Gly Leu Arg Asp Ala Ser Val Pro Ala Pro Glu Glu Ala Ala Lys
 305 310 315 320
 Gly Glu Glu Phe Asp Phe Asn Thr Pro Pro Asn Gln Gln Asn Asp Val
 325 330 335
 Lys Gly Ala Ser Arg Leu Pro Ser Pro Glu Asp Ala Ser His Pro Pro
 340 345 350
 Val Ser Arg Ala Thr Ser Arg Thr Ser His Gly Leu Gly Ser His Ser
 355 360 365
 Arg Tyr Ile Val Pro Gln Pro Arg Arg Leu Lys Ser Gly Trp Ala Gly
 370 375 380
 Gly Ile Cys Ala Thr Ser Lys Thr Arg Ile Lys Leu Arg Val Ala Thr
 385 390 395 400
 Glu Ala Glu Thr Lys Phe Asp Trp Asp Glu Pro Leu Val Leu Pro Pro
 405 410 415
 Ile Asp Pro Phe Leu Pro Pro His Thr Asp Lys Asp Ala Ala Ala Ser
 420 425 430
 Leu Val Ala Leu Tyr Arg Ala Gln Leu Thr Ser Leu Ala Glu Ala Phe
 435 440 445
 Arg Tyr Ile Arg Asp Lys Ser Phe Phe His Leu Tyr Thr Ser Phe His
 450 455 460
 Gly Thr Leu Thr Met Pro Val Gln Lys Leu Phe Ala His Pro Ser Ile
 465 470 475 480

Ala Pro Trp Ile Glu Glu Cys Asp Leu Val Leu Tyr Gln Arg Leu Thr
 485 490 495
 Lys Phe Ala Phe Thr Met Ser Leu Val Val Val Pro Asn Val Tyr Leu
 500 505 510
 Asn Arg Met Arg Ser Ile Ala Asp Arg Leu Val Pro His Ile Val Asp
 515 520 525
 Val Phe Gly Gly His Pro Glu His Val Val Gln Ala Lys Ala Gly Pro
 530 535 540
 Ala Ala Ile Phe Val Gly Ile Leu Glu Arg Met Thr Arg Val Asn Lys
 545 550 555 560
 Thr Ala His Ala Ala Ala Arg Pro Val Ala Leu Asp Ala Asn Arg Asp
 565 570 575
 Gln Met Tyr Ala Asp Trp Leu Glu Leu Val Asn Ala Arg Lys Ile Ala
 580 585 590
 Glu Cys Val Pro Thr Arg Gly Met Asp Asp Val Ala Glu Leu Leu Val
 595 600 605
 Arg Glu Met Arg Tyr Leu Val Asp Pro Lys Asn Tyr Ile Pro Asp Asp
 610 615 620
 Glu Thr Ser Asn Ala Asp Ala Gly Gly Ala Arg Ser Pro Phe Ala Val
 625 630 635 640
 Asn Arg Trp Arg Asp Phe Leu Met Ser Leu Pro Gly Lys Phe Pro Tyr
 645 650 655
 Ala Ser His Glu Asp Ile Val Trp Cys Val Glu Arg Val Gly Thr Ala
 660 665 670
 Ile Val Arg Glu Leu Thr Leu Gln Gly Gly Thr Ser Phe Thr Thr Trp
 675 680 685
 Trp Ser Ile Lys Thr Phe Leu Asp Glu Glu Ile Met Tyr Leu Ala Glu
 690 695 700
 Val Gly Gly Leu Met His Thr Lys Ser Leu Arg Asp Glu Pro Ser Arg
 705 710 715 720
 Gln Gln Ala Thr Thr Val Val Thr Glu Arg Gly Gln Lys Gly Ala Thr
 725 730 735
 Ala Asp Asp Asp Gly Glu Val Val Met Gly Asp Ala Gly Gln Glu Glu
 740 745 750
 Pro Met Ala Val Ala Ala Ser Pro Val His Thr Asp Arg Ala Pro Phe
 755 760 765
 Pro Pro Lys Gln Gly His Ala Ala Asn Met Gly Asp Ala Ala Ser Asp
 770 775 780
 Ala His Asp Asp Ser Gly Ile Gly Ile Arg Thr Pro Glu Thr Asp Phe
 785 790 795 800

65

Pro Val Asp Lys Phe Gly Leu His Glu Ala Glu Ala Gly Val Thr Glu
 805 810 815

Leu Ala Tyr Glu Gly Leu Asp Arg Glu Trp Ala Ala Pro Ala
 820 825 830

(2) INFORMATION ZU SEQ ID NO: 22:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 75 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: Peptide
- (B) LÄNGE: 1..75

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 22:

Ala Met Leu Trp Leu Asn Ser Ala Cys Val Glu Ser Lys Asn Ser Ile
 1 5 10 15

Pro Pro Gly Arg Val Tyr His Thr Tyr Ala Ser Lys Cys Thr Asp Ser
 20 25 30

Arg Val Val Val Leu Asn Pro Ala Ser Phe Gly Lys Leu Val Arg Val
 35 40 45

Val Phe Pro Ser Ile Lys Thr Arg Arg Leu Gly Val Arg Gly Glu Ser
 50 55 60

Lys Tyr His Tyr Cys Asn Phe Gln Leu Arg Asp
 65 70 75

(2) INFORMATION ZU SEQ ID NO: 23:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 76 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: Peptide
- (B) LÄNGE: 1..76

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 23:

66

Gly Glu Cys Trp Leu Lys Arg Ala Cys Glu Glu Gln Gln Asp Ala Ala
 1 5 10 15
 Val Gln Arg Asn Gln Ile Tyr Ala His Tyr Val Glu Ile Cys Asn Ser
 20 25 30
 Leu His Ile Lys Pro Leu Asn Ser Ala Ser Phe Gly Lys Leu Val Arg
 35 40 45
 Leu Leu Phe Pro Ser Ile Lys Thr Arg Arg Leu Gly Met Arg Gly His
 50 55 60
 Ser Lys Tyr His Tyr Cys Gly Ile Lys Leu Arg Gly
 65 70 75

(2) INFORMATION ZU SEQ ID NO: 24:

- (i) SEQUENZ CHARAKTERISTIKA:
- (A) LÄNGE: 76 Aminosäuren
 - (B) ART: Aminosäure
 - (C) STRANGFORM: Einzel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: Peptide
- (B) LAGE: 1..76

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 24:

Ala Leu Leu Trp Leu Met Lys Asn Cys Lys Trp Gln His Asp Ser Tyr
 1 5 10 15
 Val Pro Arg Gly Lys Ile Phe Ala Gln Tyr Ala Ser Ser Cys Ser Gln
 20 25 30
 Asn Asn Leu Lys Pro Leu Ser Gln Ala Ser Leu Gly Lys Leu Ile Arg
 35 40 45
 Thr Val Phe Pro Asp Leu Thr Thr Arg Arg Leu Gly Met Arg Gly Gln
 50 55 60
 Ser Lys Tyr His Tyr Cys Gly Leu Lys Leu Thr Val
 65 70 75

(2) INFORMATION ZU SEQ ID NO: 25:

- (i) SEQUENZ CHARAKTERISTIKA:
- (A) LÄNGE: 76 Aminosäuren
 - (B) ART: Aminosäure
 - (C) STRANGFORM: Einzel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: Peptide
(B) LÄGE: 1..76

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 25:

Thr	Val	Gln	Trp	Leu	Leu	Asp	Asn	Tyr	Glu	Thr	Ala	Glu	Gly	Val	Ser
1				5					10					15	
Leu	Pro	Arg	Ser	Thr	Leu	Tyr	Cys	His	Tyr	Leu	Leu	His	Cys	Gln	Glu
			20					25					30		
Gln	Lys	Leu	Glu	Pro	Val	Asn	Ala	Ala	Ser	Phe	Gly	Lys	Leu	Ile	Arg
		35					40					45			
Ser	Val	Phe	Met	Gly	Leu	Arg	Thr	Arg	Arg	Leu	Gly	Thr	Arg	Gly	Asn
	50						55					60			
Ser	Lys	Tyr	His	Tyr	Tyr	Gly	Leu	Arg	Ile	Lys	Ala				
65					70					75					

(2) INFORMATION ZU SEQ ID NO: 26:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 158 Aminosäuren
(B) ART: Aminosäure
(C) STRANGFORM: Einzel
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: Peptide
(B) LÄGE: 1..158

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 26:

Ile	Phe	Val	Gly	Ile	Leu	Glu	Arg	Met	Thr	Arg	Val	Asn	Lys	Thr	Ala
1				5					10					15	
His	Ala	Ala	Ala	Arg	Pro	Val	Ala	Leu	Asp	Ala	Asn	Arg	Asp	Gln	Met
			20					25					30		
Tyr	Ala	Asp	Trp	Leu	Glu	Leu	Val	Asn	Ala	Arg	Lys	Ile	Ala	Glu	Cys
		35					40					45			
Val	Pro	Thr	Arg	Gly	Met	Asp	Asp	Val	Ala	Glu	Leu	Leu	Val	Arg	Glu
		50				55					60				
Met	Arg	Tyr	Leu	Val	Asp	Pro	Lys	Asn	Tyr	Ile	Pro	Asp	Asp	Glu	Thr
65					70					75				80	
Ser	Asn	Ala	Asp	Ala	Gly	Gly	Ala	Arg	Ser	Pro	Phe	Ala	Val	Asn	Arg
			85						90					95	
Trp	Arg	Asp	Phe	Leu	Met	Ser	Leu	Pro	Gly	Lys	Phe	Pro	Tyr	Ala	Ser

68

	100		105		110
His	Glu Asp Ile Val Trp Cys Val Glu Arg Val Gly Thr Ala Ile Val				
	115		120		125
Arg	Glu Leu Thr Leu Gln Gly Gly Thr Ser Phe Thr Thr Trp Trp Ser				
	130		135		140
Ile	Lys Thr Phe Leu Asp Glu Glu Ile Met Tyr Leu Ala Glu				
145		150		155	

(2) INFORMATION ZU SEQ ID NO: 27:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 168 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: Peptide
- (B) LAGE: 1..168

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 27:

Ile	Phe	Ser	Asn	Leu	Leu	Ser	Arg	Leu	Leu	Arg	Val	Asn	Asp	Thr	Ala
1				5				10						15	
His	Ala	Ala	Ala	Arg	Phe	Leu	Ala	Asn	Pro	Ala	Asp	Arg	His	Leu	Ile
			20					25					30		
Cys	Asn	Asp	Trp	Glu	Arg	Phe	Val	Ser	Thr	Arg	Phe	Ile	Val	His	Arg
		35					40					45			
Glu	Leu	Met	Cys	Asn	Asp	Lys	Glu	Ala	Val	Ala	Ala	Leu	Asp	Glu	Trp
	50					55					60				
Tyr	Ser	Ile	Leu	Ser	Thr	Cys	Ser	Asn	Pro	Ser	Asp	Leu	Leu	Asp	Pro
65					70					75				80	
Leu	Lys	Asp	Lys	His	Glu	Ala	Ser	Asp	Thr	Ser	Met	Lys	Arg	Val	Glu
				85					90					95	
Leu	Arg	Gln	Ile	Asp	Gly	Val	Leu	Asp	Arg	Met	Ala	Asp	Phe	Phe	Leu
			100					105					110		
Glu	Leu	Pro	Ser	Arg	Phe	Pro	Ser	Cys	Ser	Pro	Arg	Met	Phe	Leu	Leu
		115						120				125			
Cys	Leu	Gly	Ala	Leu	Gln	Thr	Ser	Val	Leu	Arg	Glu	Ile	Thr	Val	Ser
	130					135						140			
Gly	Gly	Glu	Ala	Phe	Gly	Ala	Leu	Trp	Val	Ile	Arg	Cys	Trp	Val	Asp
145					150					155					160

Glu Tyr Met Thr Trp Val Ala Glu
165

(2) INFORMATION ZU SEQ ID NO: 28:

- (i) SEQUENZ CHARAKTERISTIKA:
 (A) LÄNGE: 159 Aminosäuren
 (B) ART: Aminosäure
 (C) STRANGFORM: Einzel
 (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

- (ix) MERKMALE:
 (A) NAME/SCHLÜSSEL: Peptide
 (B) LAGE: 1..159

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 28:

Asn	Phe	Thr	Lys	Leu	Val	Lys	Lys	Leu	Ile	Lys	Leu	Leu	Lys	Phe	Ile	1	5	10	15
Leu	Asn	Phe	Leu	Lys	Ser	Phe	Pro	Ile	Phe	Lys	Ser	Gly	Met	Asn	Asn	20	25	30	
Asp	Trp	Lys	Asn	Ile	Val	Asn	Leu	Asp	Asp	Ile	Leu	Glu	Met	Met	Ile	35	40	45	
Asn	Glu	Asp	Asp	Thr	Asn	Ser	Glu	Thr	Asn	Thr	Ile	Met	Gln	His	Leu	50	55	60	
Gln	Gly	Phe	Cys	Gln	Val	Phe	Val	Thr	Lys	Phe	Leu	Asn	Ser	Ser	Met	65	70	75	80
Ser	Val	Ser	Asn	Asp	Pro	Ser	Val	Ser	Ile	Glu	Cys	Lys	Ser	Leu	Asn	85	90	95	
Glu	Met	Ile	Lys	Asp	Phe	Cys	Ser	Phe	Ile	Ser	Leu	Gln	Ser	Lys	Phe	100	105	110	
Ser	Cys	Leu	Lys	Leu	Ile	Asp	Cys	Ser	Thr	Arg	Phe	Arg	Asn	Ala	Ile	115	120	125	
Ile	Gly	Asp	Ile	Ser	Leu	Lys	Ser	Asn	Glu	Asn	Leu	Leu	Ser	Trp	Leu	130	135	140	
Phe	Leu	Asn	Asn	Val	Met	Gly	Gln	Leu	Leu	Asn	Tyr	Cys	Phe	Glu	145	150	155		

(2) INFORMATION ZU SEQ ID NO: 29:

- (i) SEQUENZ CHARAKTERISTIKA:
 (A) LÄNGE: 149 Aminosäuren
 (B) ART: Aminosäure
 (C) STRANGFORM: Einzel
 (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: Peptide

(B) LAGE: 1..149

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 29:

Ala Phe Ala Gln Thr Leu Arg Arg Tyr Thr Ser Leu Asn His Leu Ala
1 5 10 15

Gln Ala Ala Arg Ala Val Leu Gln Asn Thr Ala Gln Ile Asn Gln Met
20 25 30

Leu Ser Asp Leu Asn Arg Val Asp Phe Ala Asn Val Gln Glu Gln Ala
35 40 45

Ser Trp Val Cys Arg Cys Glu Asp Arg Val Val Gln Arg Leu Glu Gln
50 55 60

Asp Phe Lys Val Thr Leu Gln Gln Gln Asn Ser Leu Glu Gln Trp Ala
65 70 75 80

Ala Trp Leu Asp Gly Val Val Ser Gln Val Leu Lys Pro Tyr Gln Gly
85 90 95

Ser Ala Gly Phe Pro Lys Ala Ala Lys Leu Phe Leu Leu Lys Trp Ser
100 105 110

Phe Tyr Ser Ser Met Val Ile Arg Asp Leu Thr Leu Arg Ser Ala Ala
115 120 125

Ser Phe Gly Ser Phe His Leu Ile Arg Leu Leu Tyr Asp Glu Tyr Met
130 135 140

Tyr Tyr Leu Ile Glu
145

(2) INFORMATION ZU SEQ ID NO: 30:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 22 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Doppel

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: exon

(B) LAGE: 1..22

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 30:

CATGAAGCTT TCGCGAGAAT TC

(2) INFORMATION ZU SEQ ID NO: 31:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 18 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..18

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 31:

GAATTCTCGC GAAAGCTT

18

(2) INFORMATION ZU SEQ ID NO: 32:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 23 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..23

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 32:

CATGCAAGCT TTCGCGAGAA TTC

23

(2) INFORMATION ZU SEQ ID NO: 33:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 19 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..19

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 33:

GAATTCTCGC GAAAGCTTG

19

(2) INFORMATION ZU SEQ ID NO: 34:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 24 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..24

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 34:

CATGCTAAGC TTTCGCGAGA ATTC

24

(2) INFORMATION ZU SEQ ID NO: 35:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 20 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..20

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 35:

GAATTCTCGC GAAAGCTTAG

20

(2) INFORMATION ZU SEQ ID NO: 36:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 44 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..44

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 36:

AAACCGTCAC CATGAAGCTT TCGCGAGAAT TCGGGTCGAT CAGG

44

(2) INFORMATION ZU SEQ ID NO: 37:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 44 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..44

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 37:

CCTGATCGAC CCGAATTCTC GCGAAAGCTT CATGGTGACG GTTT

44

(2) INFORMATION ZU SEQ ID NO: 38:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 45 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..45

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 38:

AAACCGTCAC CATGCAAGCT TTCGCGAGAA TTCGGGTCGA TCAGG

45

(2) INFORMATION ZU SEQ ID NO: 39:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 45 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..45

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 39:

CCTGATCGAC CCGAATTCTC GCGAAAGCTT GCATGGTGAC GGTTT

45

(2) INFORMATION ZU SEQ ID NO: 40:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 46 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..46

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 40:

AAACCGTCAC CATGCTAAGC TTTCGCGAGA ATTCGGGTCG ATCAGG

46

(2) INFORMATION ZU SEQ ID NO: 41:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 46 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
 - (A) NAME/SCHLÜSSEL: exon
 - (B) LAGE: 1..46

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 41:

CCTGATCGAC CCGAATTCTC GCGAAAGCTT AGCATGGTGA CGGTTT

46

(2) INFORMATION ZU SEQ ID NO: 42:

- (i) SEQUENZ CHARAKTERISTIKA:
 - (A) LÄNGE: 86 Basenpaare
 - (B) ART: Nukleinsäure
 - (C) STRANGFORM: Doppel
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..86

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 42:

AAACCGTCAC CATGCGTCTG TCCATCATCG CAGTCCTTCC CCTGGCCCTC GCCATGAAGC 60
TTTCGCGAGA ATTCGGGTCG ATCAGG 86

(2) INFORMATION ZU SEQ ID NO: 43:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 86 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..86

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 43:

CCTGATCGAC CCGAATTCTC GCGAAAGCTT CATGGCGAGG GCCAGGGGAA GGACTGCGAT 60
GATGGACAGA CGCATGGTGA CGGTTT 86

(2) INFORMATION ZU SEQ ID NO: 44:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 87 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..87

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 44:

AAACCGTCAC CATGCGTCTG TCCATCATCG CAGTCCTTCC CCTGGCCCTC GCCATGCAAG 60
CTTTCGCGAG AATTCGGGTC GATCAGG 87

(2) INFORMATION ZU SEQ ID NO: 45:

--

- (i) SEQUENZ CHARAKTERISTIKA:
(A) LÄNGE: 87 Basenpaare
(B) ART: Nukleinsäure
(C) STRANGFORM: Einzel
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
(A) NAME/SCHLÜSSEL: exon
(B) LAGE: 1..87

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 45:

CCTGATCGAC CCGAATTCTC GCGAAAGCTT GCATGGCGAG GGCCAGGGGA AGGACTGCGA 60
TGATGGACAG ACGCATGGTG ACGGTTT 87

(2) INFORMATION ZU SEQ ID NO: 46:

- (i) SEQUENZ CHARAKTERISTIKA:
(A) LÄNGE: 88 Basenpaare
(B) ART: Nukleinsäure
(C) STRANGFORM: Doppel
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
(A) NAME/SCHLÜSSEL: exon
(B) LAGE: 1..88

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 46:

AAACCGTCAC CATGCGTCTG TCCATCATCG CAGTCCTTCC CCTGGCCCTC GCCATGCTAA 60
GCTTTCGCGA GAATTCGGGT CGATCAGG 88

(2) INFORMATION ZU SEQ ID NO: 47:

- (i) SEQUENZ CHARAKTERISTIKA:
(A) LÄNGE: 88 Basenpaare
(B) ART: Nukleinsäure
(C) STRANGFORM: Doppel
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

- (ix) MERKMALE:
(A) NAME/SCHLÜSSEL: exon
(B) LAGE: 1..88

--

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 47:

CCTGATCGAC CCGAATTCTC GCGAAACGTT AGCATGGCGA GGGCCAGGGG AAGGACTGCG 60
ATGATGGACA GACGCATGGT GACGGTTT 88

(2) INFORMATION ZU SEQ ID NO: 48:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 104 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..104

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 48:

AAACCGTCAC CATGGTCACC CTCCGCCGCC TCGCCGTCCT CCTCGGCGCC ATCCCCGCCG 60
CCCTCGCCGC CATGAAGCTT TCGCGAGAAT TCGGGTCGAT CAGG 104

(2) INFORMATION ZU SEQ ID NO: 49:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 104 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..104

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 49:

CCTGATCGAC CCGAATTCTC GCGAAAGCTT CATGGCGGCG AGGGCGGCGG GGATGGCGCC 60
GAGGAGGACG GCGAGGCGGC GGAGGGTGAC CATGGTGACG GTTT 104

(2) INFORMATION ZU SEQ ID NO: 50:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 105 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: exon

(B) LAGE: 1..105

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 50:

AAACCGTCAC CATGGTCACC CTCCGCCGCC TCGCCGTCCT CCTCGGCGCC ATCCCCGCCG 60

CCCTCGCCGC CATGCAAGCT TTCGCGAGAA TTCGGGTCGA TCAGG 105

(2) INFORMATION ZU SEQ ID NO: 51:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 105 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Doppel

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: exon

(B) LAGE: 1..105

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 51:

CCTGATCGAC CCGAATTCTC GCGAAAGCTT GCATGGCGGC GAGGGCGGCG GGGATGGCGC 60

CGAGGAGGAC GGCGAGGCGG CGGAGGGTGA CCATGGTGAC GGTTT 105

(2) INFORMATION ZU SEQ ID NO: 52:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 106 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Doppel

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: exon

(B) LAGE: 1..106

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 52:

AAACCGTCAC CATGGTCACC CTCCGCCGCC TCGCCGTCCT CCTCGGCGCC ATCCCCGCCG 60

CCCTCGCCGC CATGCTAAGC TTTCGCGAGA ATTCGGGTCG ATCAGG

106

(2) INFORMATION ZU SEQ ID NO: 53:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 106 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Doppel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..106

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 53:

CCTGATCGAC CCGAAATCTC GCGAAACGTT AGCATGGCGG CGAGGGCGGC GGGGATGGCG 60

CCGAGGAGGA CGGCGAGGCG GCGGAGGGTG ACCATGGTGA CGGTTT 106

(2) INFORMATION ZU SEQ ID NO: 54

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 21 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon
- (B) LAGE: 1..21

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 54:

CCTGCTATGT ACGGACAAGG G 21

(2) INFORMATION ZU SEQ ID NO: 55:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 22 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: exon

(B) LAGE: 1..22

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 55:

CCTCGTCATG CAGGAGCCGC CC

22

(2) INFORMATION ZU SEQ ID NO: 56:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 16 Basenpaare

(B) ART: Nukleotide

(C) STRANGFORM: Einzel

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: exon

(B) LAGE: 1..16

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 56:

NNGTNRCCRG YAACNN

16

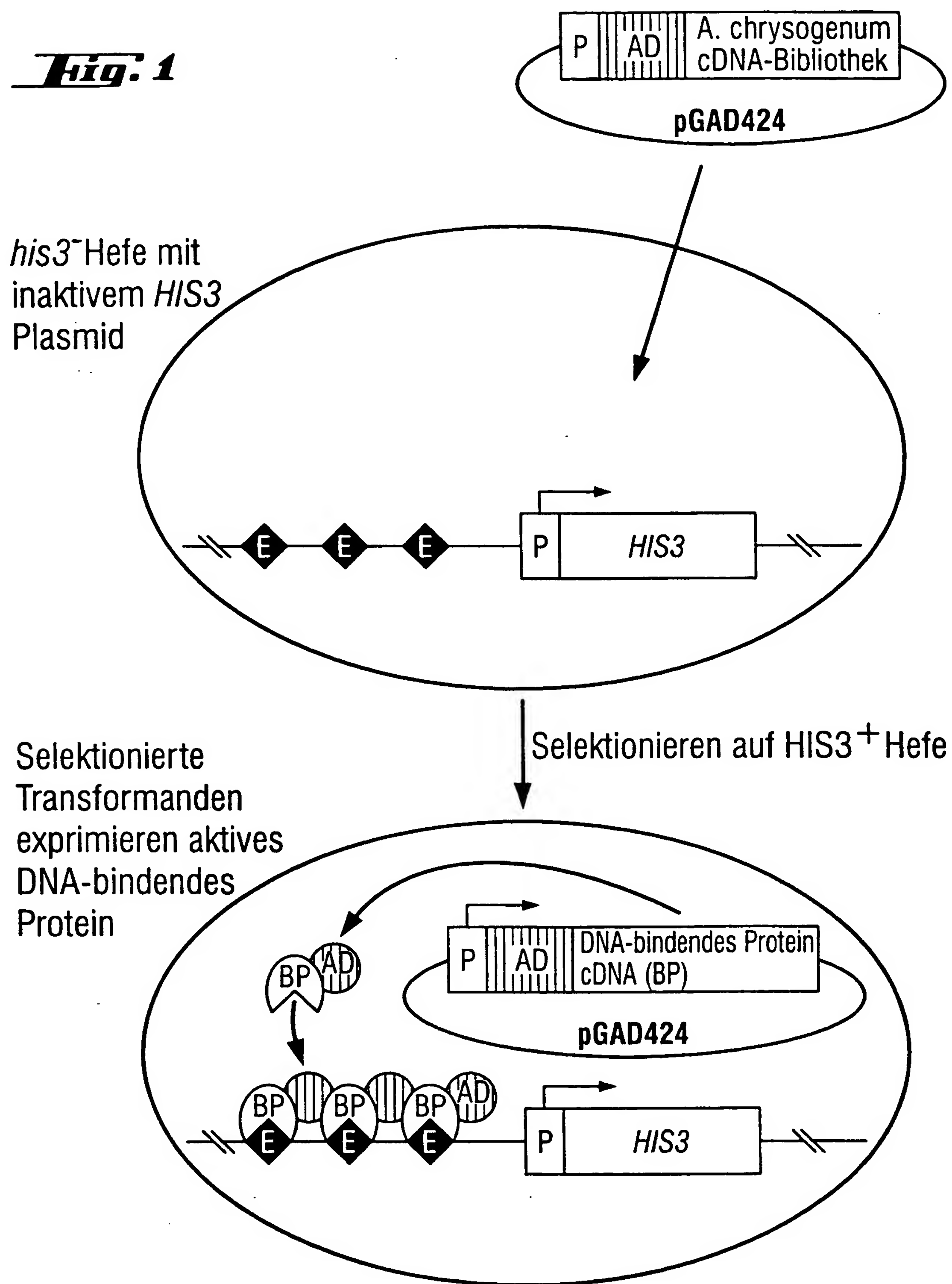
Patentansprüche:

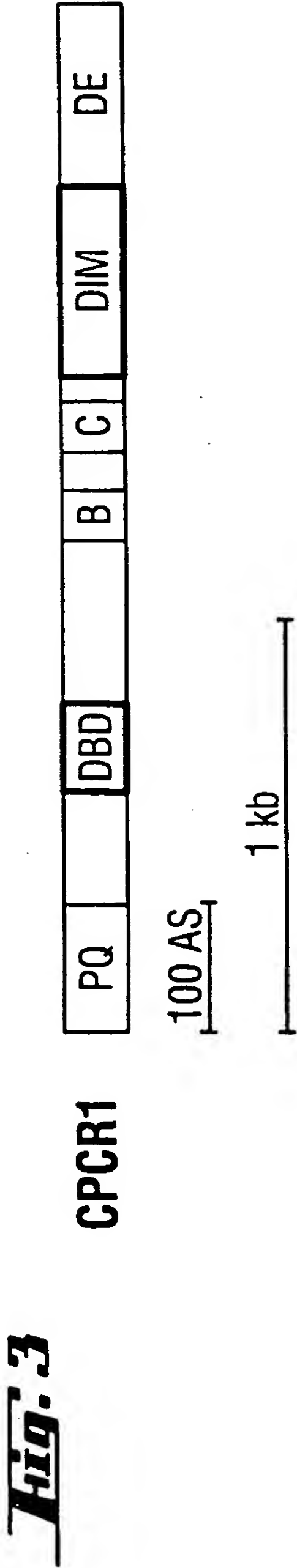
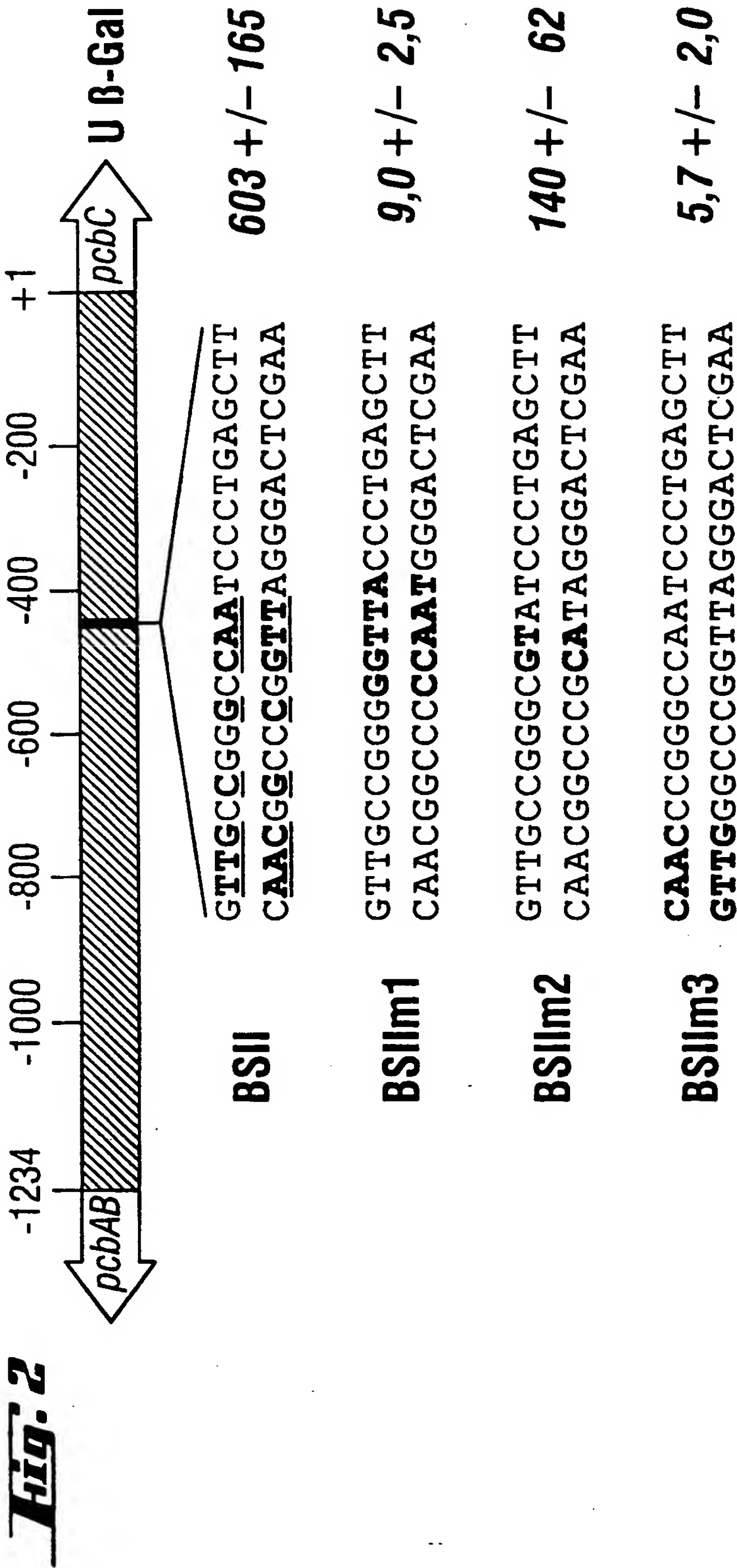
1. DNA-bindendes Protein, welches als Monomer, Homodimer oder Teil eines Heterodimer an doppelsträngige DNA der Sequenz GTTGCCGGGCCAAT CCCTGAGCTT (SEQ ID NO.: 4) aus dem intergenischen Bereich des pcbAB und pcbC-Gens von *Acremonium chrysogenum* bindet und in der Lage ist, die Transkription des pcbAB und/oder pcbC-Gens zu beeinflussen.
2. DNA-bindendes Protein gemäß Anspruch 1, dadurch gekennzeichnet, daß es die Aminosäuresequenz des CPCR1-Proteins gemäß Tab. 7 (SEQ ID NO.: 21) oder Teile davon enthält.
3. Fusionsprotein, enthaltend die Aminosäuresequenz des CPCR1-Proteins gemäß Tab. 7 (SEQ ID NO.: 21) oder Teile davon und eine Transaktivierungsdomäne, welches in der Lage ist, die Transkription des pcbAB und/oder pcbC-Gens zu beeinflussen.
4. Fusionsprotein gemäß Anspruch 3, dadurch gekennzeichnet, daß die genannten Teile des CPCR1 Proteins der DNA-Bindungsdomäne (Aminosäuren 180 bis 254 gemäß Tab. 7) und/oder der Dimerisierungsdomäne (Aminosäuren 502 bis 660 gemäß Tabelle 7) entsprechen und die Transaktivierungsdomäne vom Transkriptionsfaktor GAL4 (Aminosäuren 768 bis 881) stammt.
5. DNA-Molekül, kodierend für ein Protein gemäß einem oder mehreren der Ansprüche 1 bis 4.
6. DNA-Molekül gemäß Anspruch 5, dadurch gekennzeichnet, daß
 - (a) es eine DNA-Sequenz gemäß Tabelle 6 (SEQ ID NO.: 20) oder Teile davon enthält;

- (b) es eine DNA-Sequenz gemäß Tabelle 13 (SEQ ID NO.: 57) oder Teile davon enthält;
 - (c) es eine DNA-Sequenz enthält, die in der Lage ist, mit den DNA-Molekülen gemäß (a) oder (b) unter stringenten Bedingungen zu hybridisieren, oder Teile davon;
 - (d) es eine DNA-Sequenz enthält, die aufgrund der Degeneriertheit des genetischen Codes von den DNA-Molekülen gemäß (a), (b) und (c) verschieden ist, jedoch die Expression der entsprechend mit den DNA-Molekülen gemäß (a), (b) und (c) exprimierbaren Proteinen ermöglicht, oder Teile davon.
7. Vektor, enthaltend eine DNA gemäß Anspruch 5 oder 6.
8. Vektor gemäß Anspruch 7, dadurch gekennzeichnet, daß es ein Vektor geeignet zur Expression in *Acremonium chrysogenum* oder *Saccharomyces cerevisiae* ist.
9. Vektor gemäß Anspruch 7, dadurch gekennzeichnet, daß er ausgewählt wird aus der Gruppe enthaltend die Vektoren pGC1 und pGC1 Δ DIM1.
10. Wirtszelle, enthaltend eine oder mehrere Kopien eines Vektors gemäß einem oder mehreren der Ansprüche 7 bis 9 oder eine DNA gemäß einem der Ansprüche 5 oder 6.
11. Wirtszelle gemäß Anspruch 10, dadurch gekennzeichnet, daß sie ausgewählt ist aus der Gruppe enthaltend *Acremonium chrysogenum* und *Saccharomyces cerevisiae*.

12. Verfahren zur Herstellung eines DNA-bindenden Proteins gemäß einem oder mehreren der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß eine DNA gemäß einem der Ansprüche 5 oder 6 mittels eines Vektors gemäß einem der Ansprüche 7 bis 9 in einer Wirtszelle gemäß Anspruch 10 oder 11 exprimiert und isoliert wird.
13. Verfahren zur Steigerung der Cephalosporin C-Ausbeute in *Acremonium chrysogenum*, bei dem die Expression des cpcR1-Genprodukts mit gentechnischen Mitteln variiert wird und das Cephalosporin C isoliert wird.
14. Verfahren gemäß Anspruch 13, dadurch gekennzeichnet, daß die variierte Expression durch Einbringen einer oder mehrerer Kopien des cpcR1-Gens in den Wirtsorganismus zusätzlich zum endogenen cpcR1-Gen erreicht wird.
15. Verfahren gemäß Anspruch 13, dadurch gekennzeichnet, daß die variierte Expression durch Integration einer oder mehrerer zusätzlicher DNA-Bindesequenzen TTGCCGGGCCAAT (SEQ ID NO.: 1) des DNA-bindenden Proteins gemäß einem der Ansprüche 1 bis 4 im pcbAB/pcbC-Promotor erreicht wird.
16. DNA-Bindungsstelle für ein DNA-bindendes Protein gemäß einem oder mehreren der Ansprüche 1 bis 4, ausgewählt aus der Gruppe enthaltend die doppelsträngigen Sequenzen GTTGCCGGGCCAATCCCTGAGCTT (SEQ ID NO.: 4), TTGCCGGGCCAAT (SEQ ID NO.: 1) und GTTGCCGGGGCGTATCCCTGAGCTT (SEQ ID NO.: 8).
17. Vektor gemäß einem oder mehreren der Ansprüche 7 bis 9, enthaltend eine oder mehrere Kopien der DNA-Sequenzen gemäß Anspruch 16.

1 / 6

Fig. 1



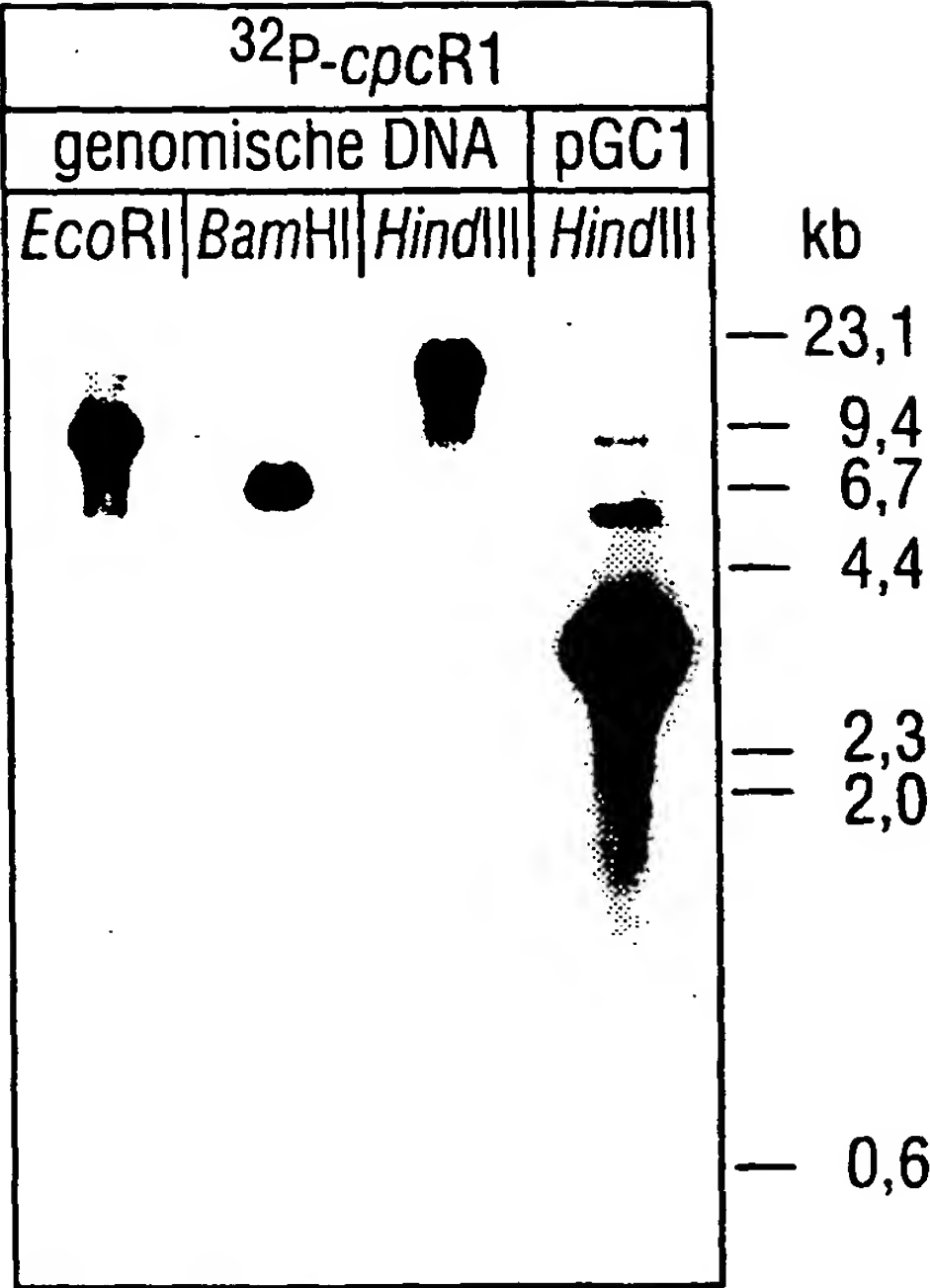


Fig. 4

4 / 6

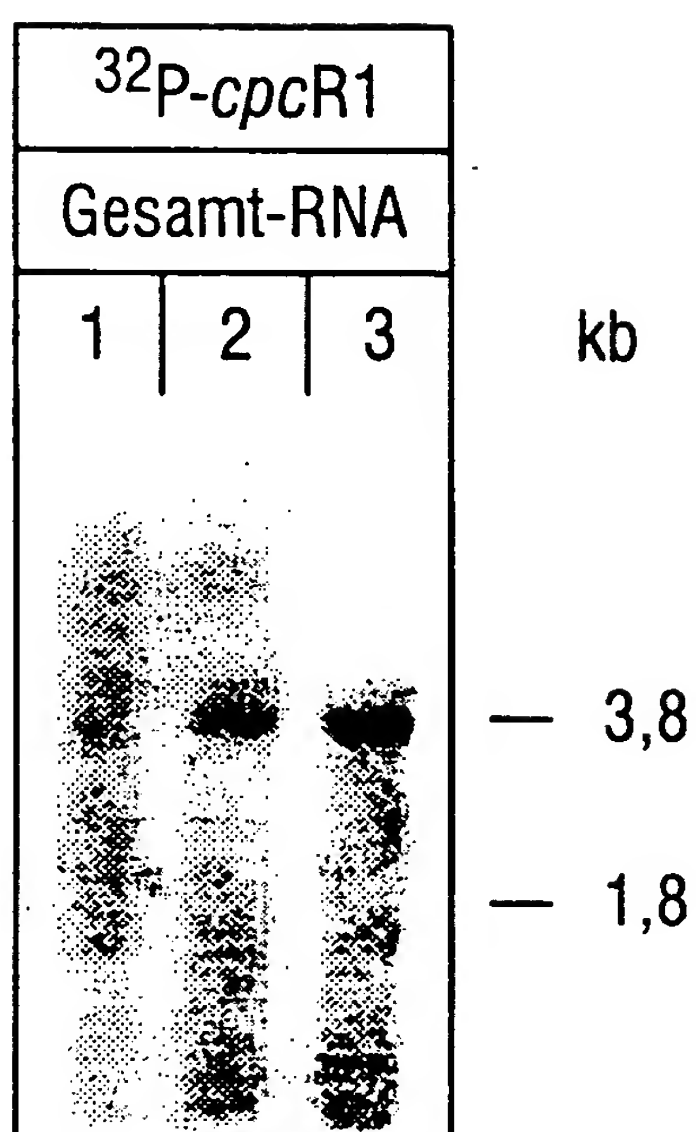
***Fig. 5***

Fig. 6

U β -Gal

603 +/– 165



AD-CPCR1

5,7 +/– 4,0



AD-DIM1

32,5 +/– 4,5



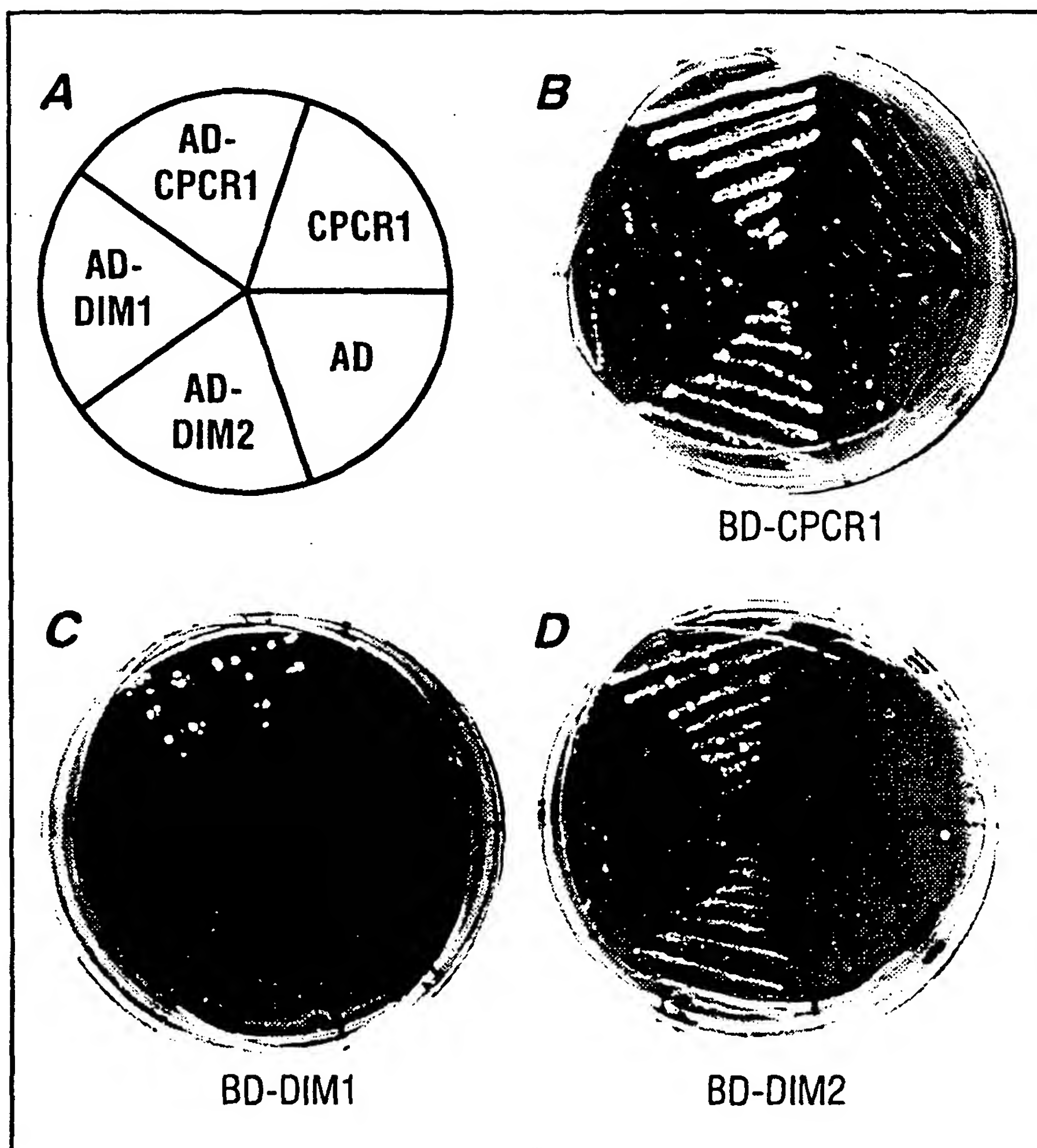
AD-DIM2

4,5 +/– 2,1



CPCR1

100 AS

***Fig. 7***

BUDAPESTER VERTRAG ÜBER DIE INTERNATIONALE
ANERKENNUNG DER HINTERLEGUNG VON MIKROORGANISMEN
FÜR DIE ZWECKE VON PATENTVERFAHREN

INTERNATIONALES FORMBLATT

Hoechst Marion Roussel

65926 Frankfurt

LEBENSFÄHIGKEITSBESCHEINIGUNG

ausgestellt gemäß Regel 10.2 von der unten angegebenen
INTERNATIONALEN HINTERLEGUNGSSTELLE

I. HINTERLEGER	II. KENNZEICHNUNG DES MIKROORGANISMUS
Name: Hoechst Marion Roussel Anschrift: 65926 Frankfurt	Von der INTERNATIONALEN HINTERLEGUNGSSTELLE zugeteilte EINGANGSNUMMER: DSM 11878 Datum der Hinterlegung oder Weiterleitung ¹ : 1997-11-28
III. LEBENSFÄHIGKEITSBESCHEINIGUNG	
Die Lebensfähigkeit des unter II genannten Mikroorganismus ist am 1997-11-28 ¹ geprüft worden. Zu diesem Zeitpunkt war der Mikroorganismus (X) ² lebensfähig () ² nicht mehr lebensfähig	
IV. BEDINGUNGEN, UNTER DENEN DIE LEBENSFÄHIGKEITSPRÜFUNG DURCHGEFÜHRT WORDEN IST	
V. INTERNATIONALE HINTERLEGUNGSSTELLE	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Anschrift: Mascheroder Weg 1b D-38124 Braunschweig	Unterschrift(en) der zur Vertretung der internationalen Hinterlegungsstelle befugten Person(en) oder des (der) von ihr ermächtigten Bediensteten: U. Wehls Datum: 1997-12-02

- ¹ Angabe des Datums der Ersthinterlegung. Wenn eine erneute Hinterlegung oder eine Weiterleitung vorgenommen worden ist, Angabe des Datums der jeweils letzten erneuten Hinterlegung oder Weiterleitung.
- ² In den in Regel 10.2 Buchstabe a Ziffer ii und iii vorgesehenen Fällen Angabe der letzten Lebensfähigkeitsprüfung.
- ³ Zutreffendes ankreuzen.
- ⁴ Ausfüllen, wenn die Angaben beantragt worden sind und wenn die Ergebnisse der Prüfung negativ waren.

BUDAPESTER VERTRAG ÜBER DIE INTERNATIONALE
ANERKENNUNG DER HINTERLEGUNG VON MIKROORGANISMEN
FÜR DIE ZWECKE VON PATENTVERFAHREN

INTERNATIONALES FORMBLATT

Hoechst Marion Roussel

65926 Frankfurt

LEBENSFÄHIGKEITSBESCHEINIGUNG

ausgestellt gemäß Regel 10.2 von der unten angegebenen
INTERNATIONALEN HINTERLEGUNGSSTELLE

I. HINTERLEGER		II. KENNZEICHNUNG DES MIKROORGANISMUS	
Name: Hoechst Marion Roussel Anschrift: 65926 Frankfurt		Von der INTERNATIONALEN HINTERLEGUNGSSTELLE zugeteilte EINGANGSNUMMER: DSM 11878 Datum der Hinterlegung oder Weiterleitung ¹ : 1997-11-28	
III. LEBENSFÄHIGKEITSBESCHEINIGUNG			
Die Lebensfähigkeit des unter II genannten Mikroorganismus ist am 1997-11-28 ² geprüft worden. Zu diesem Zeitpunkt war der Mikroorganismus (X) ³ lebensfähig () ³ nicht mehr lebensfähig			
IV. BEDINGUNGEN, UNTER DENEN DIE LEBENSFÄHIGKEITSPRÜFUNG DURCHGEFÜHRT WORDEN IST ⁴			
V. INTERNATIONALE HINTERLEGUNGSSTELLE			
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Anschrift: Mascheroder Weg 1b D-38124 Braunschweig		Unterschrift(en) der zur Vertretung der internationalen Hinterlegungsstelle befugten Person(en) oder des (der) von ihr ermächtigten Bediensteten: U. Wehls Datum: 1997-12-02	

- ¹ Angabe des Datums der Ersthinterlegung. Wenn eine erneute Hinterlegung oder eine Weiterleitung vorgenommen worden ist, Angabe des Datums der jeweils letzten erneuten Hinterlegung oder Weiterleitung.
² In den in Regel 10.2 Buchstabe a Ziffer ii und iii vorgesehenen Fällen Angabe der letzten Lebensfähigkeitsprüfung.
³ Zutreffendes ankreuzen.
⁴ Ausfüllen, wenn die Angaben beantragt worden sind und wenn die Ergebnisse der Prüfung negativ waren.


BUDAPESTER VERTRAG ÜBER DIE INTERNATIONALE
ANERKENNUNG DER HINTERLEGUNG VON MIKROORGANISMEN
FÜR DIE ZWECKE VON PATENTVERFAHREN

INTERNATIONALES FORMBLATT

Hoechst Marion Roussel

65926 Frankfurt

LEBENSFÄHIGKEITSBESCHEINIGUNG
ausgestellt gemäß Regel 10.2 von der unten angegebenen
INTERNATIONALEN HINTERLEGUNGSSTELLE

I. HINTERLEGER	II. KENNZEICHNUNG DES MIKROORGANISMUS
Name: Hoechst Marion Roussel Anschrift: 65926 Frankfurt	Von der INTERNATIONALEN HINTERLEGUNGSSTELLE zugeteilte EINGANGSNUMMER: DSM 11878 Datum der Hinterlegung oder Weiterleitung ¹ : 1997-11-28
III. LEBENSFÄHIGKEITSBESCHEINIGUNG	
Die Lebensfähigkeit des unter II genannten Mikroorganismus ist am 1997-11-28 ¹ geprüft worden. Zu diesem Zeitpunkt war der Mikroorganismus (X) ² lebensfähig () ² nicht mehr lebensfähig	
IV. BEDINGUNGEN, UNTER DENEN DIE LEBENSFÄHIGKEITSPRÜFUNG DURCHGEFÜHRT WORDEN IST ³	
V. INTERNATIONALE HINTERLEGUNGSSTELLE	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Anschrift: Mascheroder Weg 1b D-38124 Braunschweig	Unterschrift(en) der zur Vertretung der internationalen Hinterlegungsstelle befugten Person(en) oder des (der) von ihr ermächtigten Bediensteten:  Datum: 1997-12-02

- ¹ Angabe des Datums der Ersthinterlegung. Wenn eine erneute Hinterlegung oder eine Weiterleitung vorgenommen worden ist, Angabe des Datums der jeweils letzten erneuten Hinterlegung oder Weiterleitung.
² In den in Regel 10.2 Buchstabe a Ziffer ii und iii vorgesehenen Fällen Angabe der letzten Lebensfähigkeitsprüfung.
³ Zutreffendes ankreuzen.
⁴ Ausfüllen, wenn die Angaben beantragt worden sind und wenn die Ergebnisse der Prüfung negativ waren.


BUDAPESTER VERTRAG ÜBER DIE INTERNATIONALE
ERKENNUNG DER HINTERLEGUNG VON MIKROORGANISMEN
FÜR DIE ZWECKE VON PATENTVERFAHREN

INTERNATIONALES FORMBLATT

Hoechst Marion Roussel

65926 Frankfurt

EMPFANGSBESTÄTIGUNG BEI ERSTHINTERLEGUNG,
ausgestellt gemäß Regel 7.1 von der unten angegebenen
INTERNATIONALEN HINTERLEGUNGSSTELLE

I. KENNZEICHNUNG DES MIKROORGANISMUS	
Vom HINTERLEGER zugeleitetes Bezugszeichen: A 3/2	Von der INTERNATIONALEN HINTERLEGUNGSSTELLE zugeleitete EINGANGSNUMMER: DSM 11878
II. WISSENSCHAFTLICHE BESCHREIBUNG UND/ODER VORGESCHLAGENE TAXONOMISCHE BEZEICHNUNG	
Mit dem unter I. bezeichneten Mikroorganismus wurde <input type="checkbox"/> eine wissenschaftliche Beschreibung <input checked="" type="checkbox"/> eine vorgeschlagene taxonomische Bezeichnung eingereicht (Zutreffendes ankreuzen).	
III. EINGANG UND ANNAHME	
Diese internationale Hinterlegungsstelle nimmt den unter I bezeichneten Mikroorganismus an, der bei ihr am 1997-11-28 (Datum der Ersthinterlegung) eingegangen ist.	
IV. EINGANG DES ANTRAGS AUF UMWANDLUNG	
Der unter I bezeichnete Mikroorganismus ist bei dieser internationalen Hinterlegungsstelle am eingegangen (Datum der Ersthinterlegung) und ein Antrag auf Umwandlung dieser Ersthinterlegung in eine Hinterlegung gemäß Budapest Vertrag ist am eingegangen (Datum des Eingangs des Antrags auf Umwandlung).	
V. INTERNATIONALE HINTERLEGUNGSSTELLE	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Anschrift: Mascheroder Weg 1b D-38124 Braunschweig	Unterschrift(en) der zur Vertretung der internationalen Hinterlegungsstelle befugten Person(en) oder des (der) von ihr ermächtigten Bediensteten:  Datum: 1997-12-02

Falls Regel 6.4 Buchstabe d zutrifft, ist dies der Zeitpunkt, zu dem der Status einer internationalen Hinterlegungsstelle erworben worden ist.

Formblatt DSMZ-BP/4 (einzige Seite) 0196


BUDAPESTER VERTRAG ÜBER DIE INTERNATIONALE
VERKENNUNG DER HINTERLEGUNG VON MIKROORGANISMEN
FÜR DIE ZWECKE VON PATENTVERFAHREN

INTERNATIONALES FORMBLATT

Hoechst Marion Roussel

65926 Frankfurt

EMPFANGSBESTÄTIGUNG BEI ERSTHINTERLEGUNG.
ausgestellt gemäß Regel 7.1 von der unten angegebenen
INTERNATIONALEN HINTERLEGUNGSSTELLE

I. KENNZEICHNUNG DES MIKROORGANISMUS	
Vom HINTERLEGER zugeteiltes Bezugszeichen: A 3/2	Von der INTERNATIONALEN HINTERLEGUNGSSTELLE zugeteilte EINGANGSNUMMER: DSM 11878
II. WISSENSCHAFTLICHE BESCHREIBUNG UND/ODER VORGESCHLAGENE TAXONOMISCHE BEZEICHNUNG	
Mit dem unter I. bezeichneten Mikroorganismus wurde <input type="checkbox"/> eine wissenschaftliche Beschreibung <input checked="" type="checkbox"/> eine vorgeschlagene taxonomische Bezeichnung eingereicht. (Zutreffendes ankreuzen).	
III. EINGANG UND ANNAHME	
Diese internationale Hinterlegungsstelle nimmt den unter I bezeichneten Mikroorganismus an, der bei ihr am 1997-11-28 (Datum der Ersthinterlegung) ¹ eingegangen ist.	
IV. EINGANG DES ANTRAGS AUF UMWANDLUNG	
Der unter I bezeichnete Mikroorganismus ist bei dieser internationalen Hinterlegungsstelle am eingegangen (Datum der Ersthinterlegung) und ein Antrag auf Umwandlung dieser Ersthinterlegung in eine Hinterlegung gemäß Budapester Vertrag ist am eingegangen (Datum des Eingangs des Antrags auf Umwandlung).	
V. INTERNATIONALE HINTERLEGUNGSSTELLE	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Anschrift: Mascheroder Weg 1b D-38124 Braunschweig	Unterschrift(en) der zur Vertretung der internationalen Hinterlegungsstelle befugten Person(en) oder des (der) von ihr ermächtigten Bediensteten:  Datum: 1997-12-02

¹ Falls Regel 6.4 Buchstabe a zutrifft, ist dies der Zeitpunkt, zu dem der Status einer internationalen Hinterlegungsstelle erworben worden ist.


BUDAPESTER VERTRAG ÜBER DIE INTERNATIONALE
VERKENNUNG DER HINTERLEGUNG VON MIKROORGANISMEN
FÜR DIE ZWECKE VON PATENTVERFAHREN

INTERNATIONALES FORMBLATT

Hoechst Marion Roussel

65926 Frankfurt

EMPFANGSBESTÄTIGUNG BEI ERSTHINTERLEGUNG,
ausgestellt gemäß Regel 7.1 von der unten angegebenen
INTERNATIONALEN HINTERLEGUNGSSTELLE

I. KENNZEICHNUNG DES MIKROORGANISMUS	
Vom HINTERLEGER zugeteiltes Bezugszeichen: A 3/2	Von der INTERNATIONALEN HINTERLEGUNGSSTELLE zugeteilte EINGANGSNUMMER: DSM 11878
II. WISSENSCHAFTLICHE BESCHREIBUNG UND/ODER VORGESCHLAGENE TAXONOMISCHE BEZEICHNUNG	
Mit dem unter I. bezeichneten Mikroorganismus wurde <input type="checkbox"/> eine wissenschaftliche Beschreibung <input checked="" type="checkbox"/> eine vorgeschlagene taxonomische Bezeichnung eingereicht. (Zutreffendes ankreuzen).	
III. EINGANG UND ANNAHME	
Diese internationale Hinterlegungsstelle nimmt den unter I bezeichneten Mikroorganismus an, der bei ihr am 1997-11-28 (Datum der Ersthinterlegung) eingegangen ist.	
IV. EINGANG DES ANTRAGS AUF UMWANDLUNG	
Der unter I bezeichnete Mikroorganismus ist bei dieser internationalen Hinterlegungsstelle am eingegangen (Datum der Ersthinterlegung) und ein Antrag auf Umwandlung dieser Ersthinterlegung in eine Hinterlegung gemäß Budapest Vertrag ist am eingegangen (Datum des Eingangs des Antrags auf Umwandlung).	
V. INTERNATIONALE HINTERLEGUNGSSTELLE	
Name: DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH Anschrift: Mascheroder Weg 1b D-38124 Braunschweig	Unterschrift(en) der zur Vertretung der internationalen Hinterlegungsstelle befugten Person(en) oder des (der) von ihr ermächtigten Bediensteten:  Datum: 1997-12-02

Falls Regel 6.4 Buchstabe a zutrifft, ist dies der Zeitpunkt, zu dem der Status einer internationalen Hinterlegungsstelle erworben worden ist.

Formblatt DSMZ-BP/4 (einzige Seite) 0196

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/EP 98/07971

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/31 C12N15/52 C12N15/62 C12N15/80 C12N15/81
C12N1/15 C12N1/19 C07K14/37 //C12R1/75

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	THEN BERGH K. ET AL.: "Identification of a major cis-acting DNA element controlling the bidirectionally transcribed penicillin biosynthesis genes acvA (pcbAB) and ipnA (pcbC) of <i>Aspergillus nidulans</i> " JOURNAL OF BACTERIOLOGY, vol. 178, no. 13, 1996, pages 3908-3916, XP002101441 cited in the application see in particular S.3913 r.column 3.Abs. ff.; Fig.8 ---	1
A	EP 0 618 298 A (HOECHST AG) 5 October 1994 see the whole document ---	1-17
A	EP 0 200 425 A (LILLY CO ELI) 5 November 1986 see the whole document ---	1-17
-/--		

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- S* document member of the same patent family

Date of the actual completion of the international search

28 April 1999

Date of mailing of the international search report

18/05/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

KANIA, T

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 98/07971

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SMITH A W ET AL: "Regulation of isopenicillin N synthetase (IPNS) gene expression in Acremonium chrysogenum." BIO/TECHNOLOGY, (1990 MAR) 8 (3) 237-40. JOURNAL CODE: AL1. ISSN: 0733-222X., XP002101442 United States see the whole document ----	1-17
A	DATABASE WPI Section Ch, Week 9708 Derwent Publications Ltd., London, GB; Class C12, AN 97-079738 XP002101446 & ES 2 094 088 A (ANTIBIOTICOS SA) , 1 January 1997 see abstract ----	1-17
A	PAPAGIANNOPOULOS P. ET AL.: "The hapC gene of Aspergillus nidulans is involved in the expression of CCAAT-containing promoters" MOLECULAR AND GENERAL GENETICS, vol. 251, 1996, pages 412-421, XP002101443 see the whole document ----	1-17
A	CHEN H. AND KINSEY J.: "Purification of a heteromeric CCAAT binding protein from Neurospora crassa" MOLECULAR AND GENERAL GENETICS, vol. 249, no. 1, 1995, pages 301-308, XP002101444 cited in the application see the whole document ----	1-17
A	MARTIN J. AND GUTIÉRREZ S.: "Genes for beta-lactam antibiotic biosynthesis" ANTONIE VAN LEEUWENHOEK, vol. 67, 1995, pages 181-200, XP002101445 see the whole document -----	1-17

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int: ional Application No

PCT/EP 98/07971

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0618298 A	05-10-1994	CZ 9400628 A FI 941258 A HU 70266 A JP 7095890 A	15-02-1995 21-09-1994 28-09-1995 11-04-1995
EP 0200425 A	05-11-1986	AU 619596 B AU 4234789 A AU 592791 B AU 5643186 A BG 45047 A BG 45219 A CA 1309673 A DE 3683080 A DK 178386 A FI 861653 A, B, IE 58780 B JP 2059058 C JP 7087788 B JP 61254192 A SU 1780542 A SU 1623567 A US 4885251 A	30-01-1992 11-01-1990 25-01-1990 30-10-1986 15-03-1989 14-04-1989 03-11-1992 06-02-1992 23-10-1986 23-10-1986 17-11-1993 10-06-1996 27-09-1995 11-11-1986 07-12-1992 23-01-1991 05-12-1989

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 98/07971

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

IPK 6 C12N15/31 C12N15/52 C12N15/62 C12N15/80 C12N15/81
C12N1/15 C12N1/19 C07K14/37 //C12R1/75

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 6 C07K C12N

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	THEN BERGH K. ET AL.: "Identification of a major cis-acting DNA element controlling the bidirectionally transcribed penicillin biosynthesis genes acvA (pcbAB) and ipnA (pcbC) of Aspergillus nidulans" JOURNAL OF BACTERIOLOGY, Bd. 178, Nr. 13, 1996, Seiten 3908-3916, XP002101441 in der Anmeldung erwähnt siehe insbesondere S.3913 r.Spalte 3.Abs. ff.; Fig.8 ---	1
A	EP 0 618 298 A (HOECHST AG) 5. Oktober 1994 siehe das ganze Dokument ---	1-17

-/--

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen☒ Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"Z" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

28. April 1999

Absenddatum des internationalen Recherchenberichts

18/05/1999

Name und Postanschrift der Internationalen Recherchenbehörde
Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

KANIA, T

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	EP 0 200 425 A (LILLY CO ELI) 5. November 1986 siehe das ganze Dokument ---	1-17
A	SMITH A W ET AL: "Regulation of isopenicillin N synthetase (IPNS) gene expression in Acremonium chrysogenum." BIO/TECHNOLOGY, (1990 MAR) 8 (3) 237-40. JOURNAL CODE: AL1. ISSN: 0733-222X., XP002101442 United States siehe das ganze Dokument ---	1-17
A	DATABASE WPI Section Ch, Week 9708 Derwent Publications Ltd., London, GB; Class C12, AN 97-079738 XP002101446 & ES 2 094 088 A (ANTIBIOTICOS SA) , 1. Januar 1997 siehe Zusammenfassung ---	1-17
A	PAPAGIANNPOULOS P. ET AL.: "The hapC gene of Aspergillus nidulans is involved in the expression of CCAAT-containing promoters" MOLECULAR AND GENERAL GENETICS, Bd. 251, 1996, Seiten 412-421, XP002101443 siehe das ganze Dokument ---	1-17
A	CHEN H. AND KINSEY J.: "Purification of a heteromeric CCAAT binding protein from Neurospora crassa" MOLECULAR AND GENERAL GENETICS, Bd. 249, Nr. 1, 1995, Seiten 301-308, XP002101444 in der Anmeldung erwähnt siehe das ganze Dokument ---	1-17
A	MARTIN J. AND GUTIÉRREZ S.: "Genes for beta-lactam antibiotic biosynthesis" ANTONIE VAN LEEUWENHOEK, Bd. 67, 1995, Seiten 181-200, XP002101445 siehe das ganze Dokument -----	1-17

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP 98/07971

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 0618298 A	05-10-1994	CZ 9400628 A	15-02-1995
		FI 941258 A	21-09-1994
		HU 70266 A	28-09-1995
		JP 7095890 A	11-04-1995

EP 0200425 A	05-11-1986	AU 619596 B	30-01-1992
		AU 4234789 A	11-01-1990
		AU 592791 B	25-01-1990
		AU 5643186 A	30-10-1986
		BG 45047 A	15-03-1989
		BG 45219 A	14-04-1989
		CA 1309673 A	03-11-1992
		DE 3683080 A	06-02-1992
		DK 178386 A	23-10-1986
		FI 861653 A, B,	23-10-1986
		IE 58780 B	17-11-1993
		JP 2059058 C	10-06-1996
		JP 7087788 B	27-09-1995
		JP 61254192 A	11-11-1986
		SU 1780542 A	07-12-1992
		SU 1623567 A	23-01-1991
		US 4885251 A	05-12-1989
